Substrate Recognition Mechanism of Protein-only RNase P

by

Kipchumba J. Kaitany

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biological Chemistry)
in the University of Michigan
2020

Doctoral Committee:

Professor Emeritus Carol A. Fierke, Co-Chair
Assistant Professor Markos Koutmos, Co-Chair
Assistant Professor Aaron T. Frank
Assistant Professor Peter L. Freddolino
Associate Professor Patrick J. O'Brien
Associate Professor Bruce A. Palfey

Kipchumba J. Kaitany

kaitanyk@umich.edu

ORCID iD: 0000-0003-4085-5399

Acknowledgements

I would like to acknowledge my family for their continued love and support though out my education, my friends for their encouragement during graduate school and Carol Fierke and Markos Koutmos for their guidance.

Table of Contents

List of Tables

List of Figures

Abstract

Ribonuclease P (RNase P) is the enzyme responsible for catalyzing the removal of the 5' leader sequence from precursor transfer RNA (pre-tRNA) during the essential maturation process of transfer RNA (tRNA). While RNase P was first discovered in lower order organisms as an RNA-based ribozyme, recent work has revealed the existence of a protein-only RNase P (PRORP) within Eukaryotes. The existence of these two types of independently evolved RNase P enzymes, provides the rare opportunity to compare convergent evolutionary strategies of RNA- and protein-based catalysts. Previous work revealed that both the proteinaceous and RNA-based enzymes achieve nucleolytic activity through the same general two-metal ion mechanism. However, how PRORP recognizes tRNA substrates and whether this mechanism is similar to RNA-based RNase P is poorly understood.

This work investigates the substrate recognition strategy of PRORP. Mutagenesis and *in vitro* binding and catalytic activity assays identified several residues within the conserved pentatricopeptide repeat (PPR) domain of *Arabidopsis* PRORP1 which form binding interactions with pre-tRNA. A crystal structure of the PPR domain bound to tRNA is solved. Residues critical for binding are shown to form interactions with conserved regions of tRNA. This mode of RNA recognition by a PPR domain is novel, differing from the established sequence-specific RNA model. Interestingly, RNA-based RNase P uses a similar mechanism to detect tRNAs and this is additional evidence of convergent evolution between RNA and protein-only forms of RNase P. Overall this work identifies the mechanism of PPR domain pre-tRNA recognition, providing a foundation for the elucidating the function and mechanism of other PPR-motif containing proteins.

Chapter 1 Introduction

The central dogma of molecular biology provides a framework for understanding the flow of genetic information within a biological system (1). Within this dogma, biological information encoded in DNA is synthesized into mRNA through transcription, followed by synthesis of protein encoded in mRNA via translation. Critical to the process of translation is transfer RNA (tRNA), a 70-90 nucleotide RNA that serves as the physical link between mRNA and amino acids within the ribosome. tRNA is transcribed in cells as non-functional precursor tRNA (pre-tRNA) and must undergo several enzymatic maturation steps in order to play its essential role in translation (2, 3). The first of these steps is phosphodiester bond cleavage of the 5' end catalyzed by a metal-dependent endoribonuclease, Ribonuclease P (RNase P) (Figure 1-1).

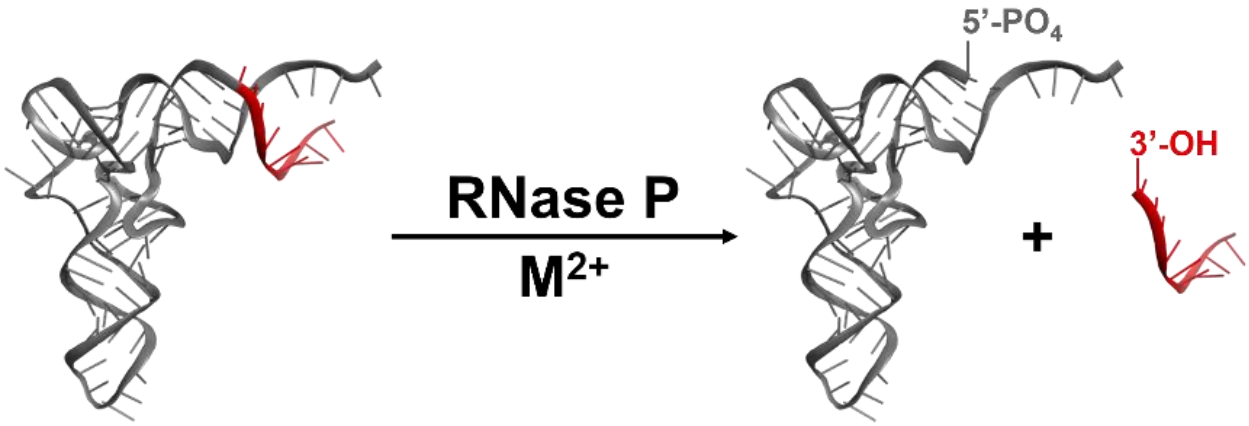


Figure 1-1 Pre-tRNA processing by RNase P.

In all domains of life, mature 5' ends of tRNA are generated by the action of ribonuclease P (RNase P). RNase P catalyzes endonucleolytic cleavage of the 5' leader sequence of precursor tRNA producing a mature 5' phosphate tRNA, and a 5' leader with a 3' hydroxyl.

1.1 Diversity of RNase P

Due to its essential role, RNase P is found in all organisms apart from one; *Nanoarchaeum equitans* is an obligate symbiont that does not require RNase P as it does not encode for pre-tRNA with 5' leaders within its genome (4, 5). Although RNase P is ubiquitous, variation in the biopolymer composition of the enzyme has led to the description of two types. The first type is a ribozyme, in which the catalytic subunit and active site are composed of RNA. The second type is protein-only RNase P, where the catalytic subunit is composed of protein and contains no RNA components. The existence of these two types of independently evolved RNase P provide the rare opportunity to compare convergent evolutionary strategies of RNA and protein-based catalysts.

1.1.1 Ribonucleoprotein (RNP) RNase P

The discovery that tRNAs within *E. coli* are transcribed as precursor tRNAs containing elongated 5' and 3' ends, led to a search for tRNA processing nucleases that revealed the first RNase P enzyme (6). The discovery of a 5' endonuclease composed of a ~400nt catalytic RNA and ~120 amino acid protein subunit was the first example of an RNA catalyst (7, 8). This momentous work was carried out by Sidney Altman and colleagues; they shared the 1989 noble prize in chemistry with Thomas Cech for their "discovery of catalytic properties of RNA" (9). Bacterial RNase P remains the most well characterized RNase P enzyme due to it serving as a simple model for studying RNA catalysis.

The bacterial RNase P RNA subunit, P RNA, alone is capable of efficient pre-tRNA processing *in vitro* in high ionic and divalent metal ion concentrations (8). The bacterial RNase P protein component, RnpA, is required for efficient catalytic activity under cellular conditions (10,

11). The protein component directly contacts the 5' leader of pre-tRNA and enhances the binding affinity for pre-tRNA and divalent metal ions (12).

Subsequent searches for RNase P activity in varying organisms revealed that RNA-based RNase P is present across all three domains of life (13). In all cases, RNA-based RNase P enzymes are ribonucleoprotein (RNP) complexes containing one catalytic RNA subunit in complex with up to ten protein subunits depending on the organism. Prokaryotic RNase P tends to contain fewer protein subunits than eukaryotic RNPs, potentially implicating a greater need for protein-based characteristics, such as pH and thermal stability, within eukaryotes (14). Human and yeast nuclear RNase P are the most extensively studied eukaryotic RNA-based RNase P. In *Saccharomyces cerevisiae* the RNP complex is composed of one RNA and nine protein subunits (15). Biochemical and mass spectrometric characterization of this complex *in vitro* revealed that all ten subunits are required for optimal activity. The additional protein subunits produce a 20-fold increase in protein content compared to bacterial enzymes, consistent with a greater functional complexity of eukaryotic RNase P as these enzymes have been shown to catalyze cleavage of multiple substrates besides pre-tRNA (15–17).

1.1.2 Eukaryotic Protein-only RNase P (PRORP)

The first indication of a protein-only RNase P came from the discovery of RNase P activity in eukaryotic organelles that exhibited biochemical characteristics that differ from the RNA-based enzyme (18). The initial substantial indications of a protein-based RNase P came from human mitochondrial RNase P activity displaying insensitivity to nuclease treatment and a density more consistent with proteins than RNA (18). These novel characteristics led Rossmanith and colleagues to perform an extensive search for the essential genetic components of human mitochondrial RNase P which in turn produced the discovery of the first protein-only RNase P

(19). Human mitochondrial RNase P was identified as a complex of three protein subunits: a tRNA m1G/C^9 methyltransferase (MRPP1/TRMT10C), a hydroxysteroid 17-β dehydrogenase 10 (MRPP2/SDR5C1), and a metallonuclease (MRPP3/*Hu*PRORP) (20). MRPP3 contained a C-terminal domain with strong homology to the flap endonucleases, a N-terminal pentatricopeptide repeat (PPR) domain (20). Discovery of human PRORP (huPRORP/MRPP3) enabled the genetic identification of MRPP3 homologs within other eukaryotic organisms such as plants, algae, and protists.

PRORP activity was first demonstrated in three *Arabidopsis thaliana* PRORP isoforms: *At*PRORP1, *At*PRORP2, and *At*PRORP3 (21). Although all three isoforms are encoded in nuclear genome, *At*PRORP2 and *At*PRORP3 localize to the nucleus, while *At*PRORP1 localizes to the mitochondria and chloroplast. *At*PRORP2 and *At*PRORP3 are thought to be redundant as they share substrate selectivity and knockout of one isoform does not eliminate RNase P activity while double deletion of both is nonviable (22). This observation, in addition to the lack of RNA RNase P homologs in plant genomes, has led to the proposal that *A. thaliana* does not encode an RNA-based RNase P. In contrast to the metazoan MRPP3, biochemical analysis of the *A. thaliana* PRORP homologs demonstrated that these enzymes function as single protein subunits and do not require additional subunits for activity (21).

Like *A. thaliana*, the model bryophyte organism *Physcomitrella patens* (spreading earthmoss) expresses three PRORP homologs and does not appear to encode RNA-dependent RNase P (23). In *P. patens*, two PRORP enzymes, *Pp*PPR67 and *Pp*PPR104 (pentatricopeptide repeat protein 67 and 104), localize to the mitochondria and chloroplast (23). The third PRORP variant, *Pp*PPR63, localizes to the nucleus. Interestingly, knockout of *Pp*PPR63 is nonlethal which may be indicative of an additional source of nuclear RNase P activity (23). The putative

mitochondrial PRORP, *Pp*PPR67, may also localize to and function in the nucleus since it contains a nuclear localization sequence (NLS) identical to the one encoded in *Pp*PPR63. In localization studies, *Pp*PPR67 was labeled through N-terminal GFP fusion downstream of the NLS (23). Consequently, if alternative splicing of *Pp*PPR67 leads to nuclear localization it would not have been observed in the localization studies.

The alga, *Ostreococcus tauri*, encompasses a strikingly diverse inventory of RNase P variants. The *O. tauri* nuclear genome encodes for both a bacterial-like RNase P protein and an *A. thaliana* PRORP1 homolog, while the bacterial-like P RNA is encoded in the mitochondria and chloroplast genomes (24). Recombinant *Ot*PRORP catalyzes 5' pre-tRNA processing *in vitro*. The bacterial-like P protein activates bacterial P RNA, however, *in vitro* transcribed organellar P RNA of *O. tauri* exhibited no RNase P activity. Additionally, attempts to reconstitute *O. tauri* P RNA with either bacterial or *O. tauri* P protein did not produce activity. This may indicate that similarly to all other eukaryotic RNA-based forms, *O. tauri* P RNA requires additional protein subunits for traditional RNase P activity or functions in a task distinct from catalyzing 5' tRNA maturation (24). Further studies are needed to identify subcellular localization and function of these proteins and RNA *in vivo*.

In the protozoan parasite *Trypanosoma brucei*, the causative agent of African sleeping sickness, all pre-tRNA 5' processing is catalyzed by two PRORP isoforms. One isoform, *Tb*PRORP1, localizes to the nucleus while *Tb*PRORP2 localizes to the mitochondria (25, 26). Both PRORP homologs exhibit RNase P activity *in* vitro. Interestingly however, *T. brucei* and other trypanosomatids mitochondria do not encode tRNAs, but instead import them from the cytosol. Moreover, *in vivo* studies suggest these imported tRNAs are already processed and mature (26). This leaves the functional role of *Tb*PRORP2 a question.

1.1.3 Homologs of *Aquifex* RNase P (HARP)

At the time the work in the following chapters was carried out, protein-only forms of RNase P were thought to be unique to eukaryotic organisms (27). The discovery of PRORP within mammalian organelles led to subsequent genetic identification of PRORP homologs within many eukaryotic organisms. This belief supported an evolutionary model in which RNA-based RNase P had been replaced by proteinaceous RNase P within eukaryotes. Yet examples of early transitional proteinaceous enzymes displaying RNase P activity had not been found. In 2017, activity-based purification of RNase P in the hyperthermophilic bacterium *Aquifex aeolicus* yielded the discovery of a new form of RNA-free RNase P, termed Homologs of *Aquifex* RNase P (HARP) (28). Thereby potentially elucidating the evolutionary missing link in RNase P evolution.

The identified enzyme is a 23-kDa protein (Aq_880) and the smallest known form of RNase P. Bioinformatic searches revealed that Aq_880 homologs are encoded in many archaea and several bacteria, however none in Eukarya (28). Due to these findings, it is proposed that bacteria acquired this enzyme through horizontal gene transfer from archaea. Interestingly, plasmid expression of Aq_880 was able to rescue lethal RNase P mutant strains in both *S. cerevisiae* and *E. coli*, although in both cases growth rate was reduced (28).

Although HARP was initially discovered and characterized in bacteria, it is primarily found within archaea. Interestingly, studies on the archaea *Haloferax volcanii* and *Methanosarcina mazei* showed knockout of their HARP homologs had no effect on growth, while knockdown or knockout of *H. volcanii* RNP RNase P resulted in severe tRNA maturation defects and lethality, respectively (29). Therefore, HARP may not be a major source of 5’ tRNA maturation within archaea, but instead serve some unidentified function in RNA processing.

1.2 Structure of RNase P

1.2.1 Structure of RNP RNase P

The bacterial RNP RNase P is composed of a large (350-420 nt) highly conserved catalytic P RNA and a small (14 kDa) protein subunit (30). The P RNA contains two individually folding domains, a catalytic (C) and a specificity (S) domain, which contain 5 conserved regions (CRI-V) found in all RNA-based RNase P (31). The C-domain accounts for about 60% of P RNA and contains CR-I, IV and V. The S-domain comprises 40% of P RNA and contains the remaining sequence conserved regions, CR-II and III (30, 32).

Structural studies utilizing a large variety of techniques have been carried out on a number of bacterial RNase P enzymes (30, 33–38). Together these works have provided insight into the structure-function relationships at play in the bacterial ribozyme and informed further mechanistic and evolutionary study. One of the more informative of these studies was the X-ray crystal structure of the *Thermatoga maritima* holoenzyme bound to $tRNA^{phe}$ and a 5' leader product (35). The structure confirmed the importance of conserved regions within the C-domain for active site formation where CR-I, IV and V enter non-helical conformations in which highly conserved residues interact with the tRNA acceptor stem, 5' leader, P protein and likely two divalent metal ions. While nucleotides in CR-II and III of the S-domain intercalated between bases in the D and TΨC loops forming base-stacking interactions with the tRNA elbow.

Although archaeal and eukaryotic P RNA is believed to fold similarly to the bacterial RNA, atomic resolution structures of these P RNAs have not been solved. Many of the regions shown to be important for substrate recognition and especially enzyme catalysis in bacterial P RNA are predicted to be structurally conserved. However, despite the presence of conserved regions and similarities in secondary structure to bacterial counterparts, eukaryotic and archaeal P RNA lack

several tertiary structural elements that stabilize bacterial P RNA (39). Absence of these structural elements may be reflected in the increased presence of protein subunits within archaeal and eukaryotic RNA-based RNase P which require at least 4 and 9 proteins, respectively.

1.2.2 Structure of Eukaryotic PRORP

X-ray crystal structures of plant and metazoan PRORP have been solved, revealing similar folds and overall structural homology within eukaryotic protein-based RNase P (40–43). The first reported PRORP structure was the crystal structure of functional recombinant *At*PRORP1 (40). This construct (Δ76PRORP1) contained a 76 amino acid N-terminal truncation removing a mitochondrial targeting signal sequence. In this structure, PRORP1 adopts an overall "V"-shaped topology, in which the N-terminal arm, a PPR RNA-binding domain, is linked to the C-terminal arm, a novel Nedd4-bp1, YacP nuclease (NYN) metallonuclease domain, through a central Zn^{2+}-binding domain (Figure 1-2) (40). The pentatricopeptide repeat (PPR) domain contains 5.5 tandem PPR motifs, wherein each motif is composed of 35 amino acids which form two anti-parallel α-helices that interact in a helix-turn-helix conformation. This series of PPR motifs form a superhelical structure containing a groove characteristic of other PPR RNA binding proteins. The PPR domain is linked to the metallonuclease domain through a conserved bipartite structural Zn^{2+} domain where a Zn^{2+} ion is coordinated between two anti-parallel β-sheets and two extended loops. The NYN metallonuclease domain contained an active site occupied by two Mn^{2+} ions which are coordinated by four conserved aspartate residues (D399, D474, D475, and D493) (40).The structure of *At*PRORP2 was also solved, sharing a similar overall "V" topology with *At*PRORP1 but with several differences (41). First, the X-ray crystal structure of *At*PRORP2 folded into a more open "V" conformation, where the PPR and metallonuclease domain form a larger degree angle.

Additionally, no metal ions were bound in the *At*PRORP2 active site, where conserved aspartate residues, which coordinated Mn^{2+} ions in *At*PRORP1, formed oligomeric interactions with the PPR domain of adjacent subunits. This oligomerization is thought to be a nonfunctional artifact of crystallization; however, it may contribute to the larger angle observed between the PPR and metallonuclease domain.

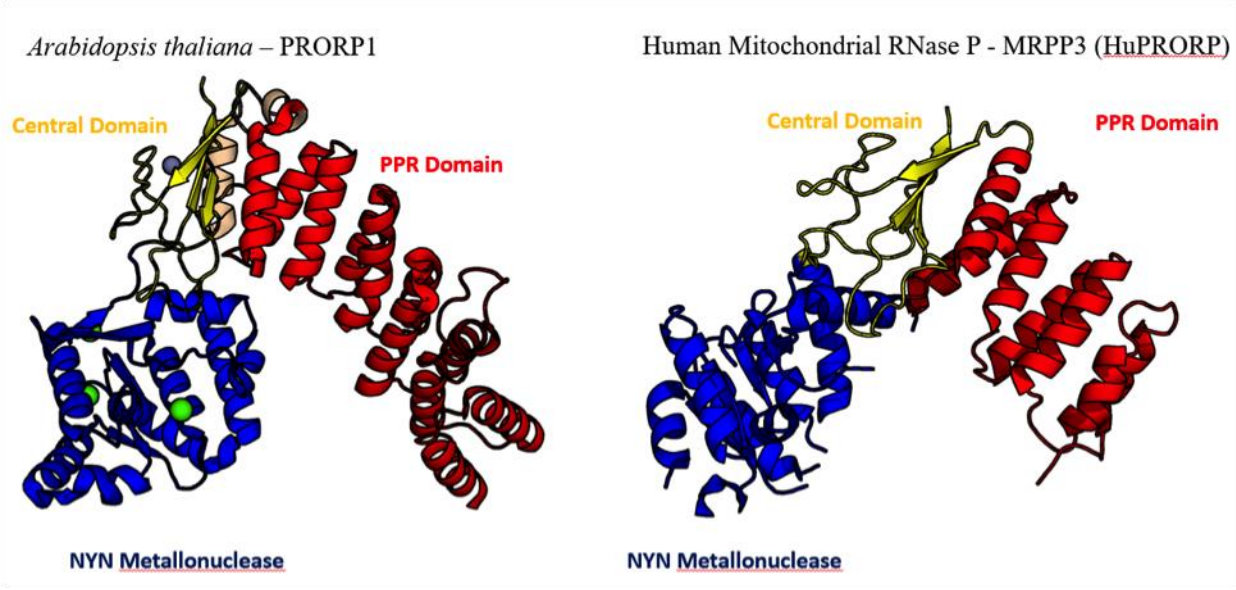


Figure 1-2 Structural similarities and differences between AtPRORP1 and HuPRORP
The overall topology and domain positioning of A. thaliana PRORP1 **(A)** is shared with huPRORP/MRPP3 **(B)**.

Structures of all three proteins that constitute human mitochondrial RNase P have been solved. MRPP1/TRMT10C is predicted to fold into two distinct domains, an N-terminal protein binding domain and a C-terminal SPOUT (SpoU and TrmD) methyltransferase domain (44). Although a full-length structure of MRPP1 is not known, an X-ray crystal structure of the C-terminal methyltransferase domain has recently been solved (45). This domain contains a canonical SPOUT α/β fold with a *S*-adenosyl-L-methionine (SAM) bound to the methyltransferase active site. The reported structure of MRPP2/SDR5C1 revealed subunits forming a tetramer like

those formed by other short-chain dehydrogenase/reductase (SDR) proteins, with each subunit in direct contact with the other three subunits. Two crystal structures have been solved of MRPP3/*Hu*PRORP, revealing an overall "V"-shaped topology analogous to the *At*PRORP isoforms (Figure 1-2) (42, 43). Both structures were of recombinant constructs containing significant N-terminal truncations of the PPR domain and exhibited disordered active sites lacking metal ions in the NYN metallonuclease active site. While the authors propose this deviation from the single-subunit PRORP active site is due to a need for the additional protein subunits, addition of MRPP1/2 to this construct *in vitro* did not display activity. Therefore, the mechanism of activation of MRPP3 by MRPP1/2 is still unclear.

Although no atomic resolution structure of the human mitochondrial RNase P holoenzyme exists, small angle X-ray scattering produced low-resolution models of the ternary complex (45). The *ab initio* models supported previous investigations that proposed a stoichiometry of 2 MRPP1, 4 MRPP2 and 1 MRPP3 to 1 pre-tRNA, although the complex stoichiometry remains a question. Other aspects of these generated models are also unclear. The positioning of pre-tRNA within the ternary complex seems functionally counterintuitive, as it makes no contact with MRPP3 in either the PPR or metallonuclease domain. Additionally, a direct interaction of MRPP2 and MRPP3 seemingly without MRPP1 or pre-tRNA involvement conflicts with size exclusion chromatography (SEC) and *in vitro* pull-down observations which saw no evidence of MRPP2-MRPP3 complexation (28, 45, 46).

1.2.3 Structure of HARP

To date no structures of HARP enzymes have been reported. However, insights based on sequence homology and SEC data have provided insights on the ribonuclease structure and stoichiometry of HARP. Bioinformatic investigation of Aq_880 predicts a PIN domain-like fold

similar to the characteristic NYN metallonuclease domains found in PRORP (28). Primary structure alignment of this putative domain with *At*PRORP metallonuclease domain identified three aspartate residues with predicted functional equivalency to essential aspartates of the PRORP active site (28). Individual alanine mutation of these conserved aspartates abolished activity *in vitro*. These findings support the presence of a metallonuclease domain with similar structure to the NYN domain found in PRORP. Observations of a ~70 kDa protein complex of native and recombinant Aq_880 (~23 kDa) were observed through SEC and mass spectrometry, suggesting the formation of a stable homotrimer (28).

1.3 RNase P Catalysis

1.3.1 Catalysis by Bacterial RNase P

As the first discovered true ribozyme, the elucidation of the chemical mechanism of RNase P was highly sought as an example of RNA-based catalytic strategies. Early RNase P processing studies demonstrated that *in vitro* and *in vivo* cleavage by RNase P resulted in 5'-phosphate and 3'-hydroxyl products (47). Additionally, no evidence of transient products or covalent enzyme–substrate bond formation was found. Catalytic studies on the RNA subunits of *E. coli* (M1 RNA), and *B. subtilis* (P RNA) RNase P, demonstrated that catalytic activity required divalent metal ions (8, 47). The results showed that both Mg^{2+} and Mn^{2+} metal ions activated endonuclease activity in a concentration dependent manner. However, Mn^{2+} was only active in the presence of Mg^{2+}, while Mg^{2+} alone displayed the greatest catalytic efficiency (48). Due to this, Mg^{2+} was deemed the putative native cofactor of RNase P. Additionally, catalytically inactive divalent metal ions, such as Ca^{2+}, inhibit activity through competition with Mg^{2+} binding sites. Based on these initial findings, Altmann and colleagues hypothesized that metal ion dependence was due to the functional role of two RNA bound divalent cations (49). One of these being a catalytic Mg^{2+} ion

directly involved in chemistry, which forms a Mg^{2+}-H_2O to catalyze nucleophilic hydrolysis of the phosphodiester bond, while a second divalent metal ion fulfills a more structural role e.g., coordinating functional groups or facilitating conformational change. Decades later, subsequent studies demonstrated that a minimum of two metal ions were important for the catalytic activity while additional metal ions were important for stabilizing the structure of P RNA and pre-tRNA (10, 11, 50–52). These data have led to the proposed chemical mechanism shown in Figure 1-3 (35).

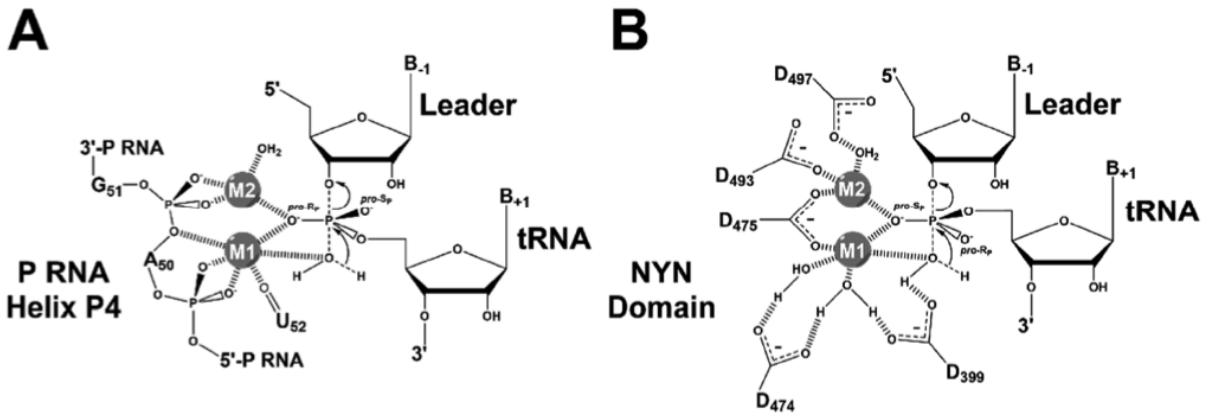


Figure 1-3 A catalytic mechanism of RNase P

Both the bacterial **(A)** RNA-dependent enzyme and PRORP **(B)**, share a general two metal ion mechanism.

A minimal four-step kinetic mechanism has been proposed for *B. subtilis* RNase P (Figure 1-4) (11). This mechanism consists of RNase P (E) and pre-tRNA (S) association (ES) through a bimolecular binding event followed by a conformational change to a catalytically competent conformer (ES*). The catalytically competent conformation then proceeds to product formation of 5’ mature tRNA (P) and a 5’ leader (L) through phosphodiester bond cleavage followed by product release. All of the kinetic steps leading to the formation of products are metal ion dependent. Chemical cleavage of pre-tRNA exhibits a log-linear dependence on pH through a Hill coefficient

of n_H = 1. This relationship agrees with the proposed chemical mechanism in which a metal hydroxide acts as the nucleophile in catalysis.

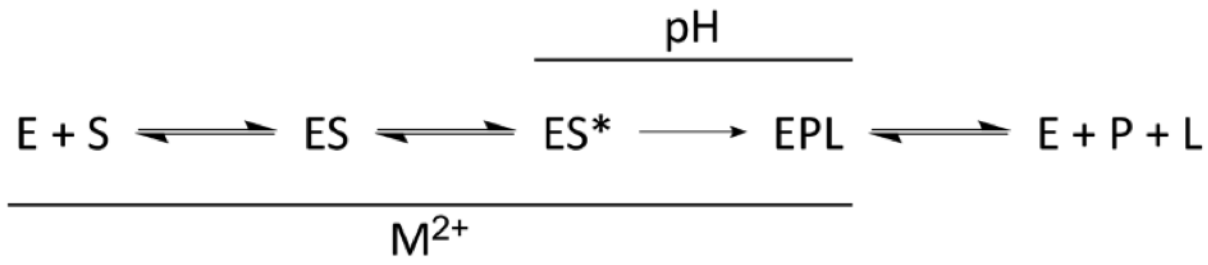


Figure 1-4 Minimal kinetic mechanism of bacterial RNase P

Bimolecular binding followed by a conformational change lead to pH-dependent catalysis and product release.

1.3.2 Catalysis by PRORP

One of the proposed theories within the RNA world theory is that RNA catalysts were replaced by protein enzymes through selectivity for increased catalytic efficiency. This does not seem to be the case in the evolution of RNase P from an RNA to a protein catalyst, where kinetic studies have shown that the efficiency of pre-tRNA cleavage is equivalent or reduced. Interestingly, this comparable catalytic efficiency may be partially due to a shared chemical mechanism (53). The elucidation of the structure of *At*PRORP1 revealed an active site containing two divalent metal ions analogous to those found in the structure of bacterial RNase P (40). Additionally, the catalytic activity of PRORP had a cooperative dependence on magnesium concentration (n = 2) and mutation of the aspartate residues coordinating these ions led to drastic decreases in activity. Furthermore, the pre-tRNA cleavage activity of PRORP revealed a single ionization with enhanced activity a high pH (pK_a = 8.7), consistent with deprotonation of a metal-bound water nucleophile (53). These data have led to a putative catalytic mechanism thought to be shared amongst all protein-only RNase P enzymes (Figure 1-3) (53). Overall, the similarities

between this chemical mechanism and that of the RNA-based enzyme support a catalytic mechanism in which either the protein or RNA produce a scaffold that binds substrate and facilitates divalent metal ion hydrolysis.

1.4 Substrates and Substrate Recognition

RNase P enzymes recognize a diverse set of substrates and catalyzes the cleavage of several RNA substrates *in vivo* and *in vitro*. Although the essential role of RNase P in tRNA processing remains the most studied, bacterial RNase P has shown to have similar catalytic efficiencies for several non-tRNA substrates: pre-4.5S RNA, pre-tmRNA, as well as various mRNAs and riboswitches (54). Additionally, noncoding RNAs have been shown to be native substrates of *S. cerevisiae* RNA-dependent RNase P (55). Within plants, the ability of PRORPs to efficiently recognize and cleave structured viral RNA has been utilized to engineer virus resistant plants. This diverse substrate pool implicates the importance of RNase P in cellular RNA maintenance. However, pre-tRNA is the only RNase P substrate that is common among all organisms, and tRNA recognition and processing remains the most studied functional role of this enzyme. As such, pre-tRNA will remain the focus of my discussion of RNase P substrate recognition.

1.4.1 Structural Features of Pre-tRNA

The essential role that RNase P plays in processing of tRNA presents the challenge of recognizing a specific set of diverse RNA substrates with varying sequences. Due to the diversity of tRNA sequences, it is thought that a major aspect of RNase P substrate specificity is recognition of conserved pre-tRNA structural features. Throughout all domains of life most tRNAs share a canonical structure composed of a cloverleaf secondary structure which folds into a L-shaped tertiary structure (Figure 1-5) (56) . However, one of the few examples where tRNAs stray from this structural canon is within mitochondrial-encoded tRNAs of the Animalia kingdom. Within

animal mitochondria tRNAs are divided into 4 types based on their primary and secondary structure: Type 0 is similar to the canonical tRNA preserving canonical intermolecular interactions and fold, Type I and II contain truncated structural elements that result in deviations from conserved tertiary interactions, while Type III lacks the conserved D-loop entirely (Figure 1-5) (57). This deviation from canonical structure has led to the suggestion that the additional protein subunits of metazoan mtRNase P may have evolved specifically to recognize these substrates.

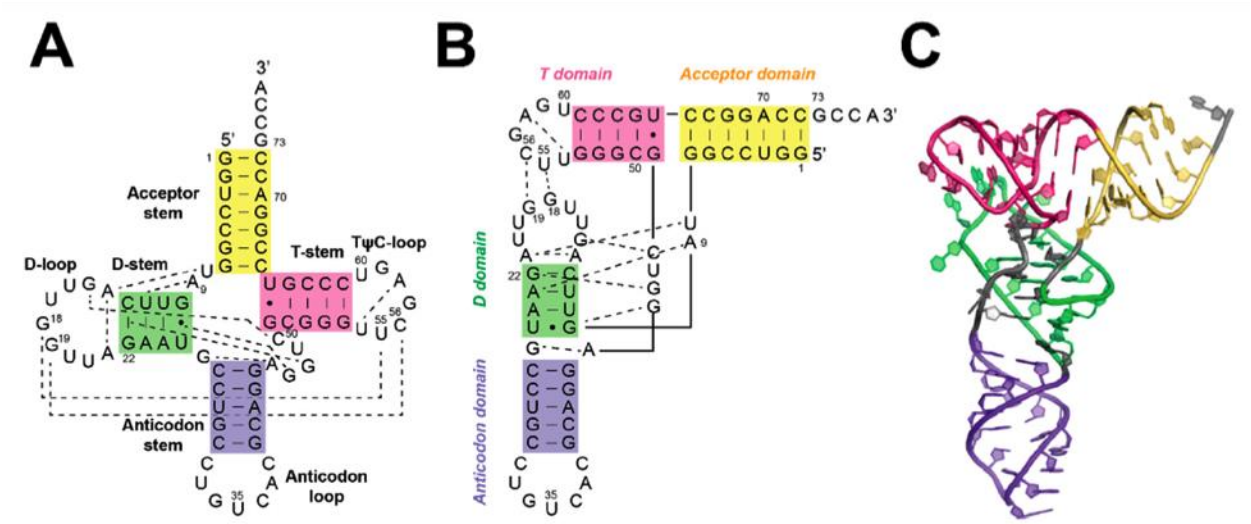


Figure 1-5 Structural features of tRNAs [adapted from Xin Liu]
The **(A)** primary and **(B)** secondary structure of *B. subtilis* $tRNA^{Asp}$, highlighting conserved tertiary regions and interactions. **(C)** The overall tertiary fold of tRNA.

1.4.2 Substrate Recognition by Bacterial RNase P

As mentioned previously, the protein component of bacterial RNase P, RnpA, is not necessary for catalysis *in vitro*. Studies investigating the role of RnpA in *B. subtilis* RNase P revealed that although the protein subunit plays modest roles in product dissociation and phosphodiester bond cleavage, it predominantly affects pre-tRNA affinity (58, 59). Compared to P RNA alone, the *Bs*RNase P holoenzyme exhibits a 10-fold greater affinity for pre-$tRNA^{Asp}$ *in vitro* (60). This increase in substrate affinity is achieved through direct interactions with the 5'

leader sequence of pre-tRNA (51). Kinetic and binding studies on varied 5' leader lengths showed that while the cleavage rate constant was unaffected, substrate binding affinity for the holoenzyme, but not P RNA alone, varied with the leader length. Specifically, it was shown that binding affinity for the holoenzyme increased with leader length from 2 to 5 nucleotides, indicating that interactions between these 5' leader nucleotides and RnpA are important for substrate recognition (61). Interactions between the 5' leader and RnpA subunit also increase magnesium ion affinity for the RNase P-pre-tRNA complex.

Elements of sequence-specificity also contribute to bacterial RNase P pre-tRNA recognition. In *E. coli* and *B. subtilis*, analysis of pre-tRNA gene sequences revealed consensus recognition elements within the 5' leader and the 3' RCCA motif (62). Sequence preference in the 5' leader is predominately seen in the -1 and -2 positions, with significant preference for uracil in the -1 position, and adenine in the -2 position (61). This base-specific preference is supported by the statistical partiality for these bases within pre-tRNA substrates. At the 3' end, the tRNA RCCA motif base pairs to a GGU sequence within the C-domain, increasing substrate and Mg^{2+} ion affinity.

Interactions of RNase P with the tRNA D-/T-loop domain play an important role in substrate recognition and cleavage. The tRNA D-/T-loop region is a unique tRNA structural motif that is utilized as a site of tRNA recognition by several enzymes. Prior to the solution of the *T. maritima* RNase P-tRNA structure, multiple studies using a variety of techniques had indicated that residues of the tRNA D-/T-loop region in pre-tRNA interact with P RNA upon binding (63–70). Studies investigating the importance of the D-/T-loop for substrate recognition by bacterial RNase P demonstrated that various model substrates which mimicked the tRNA D-/T-loop region and acceptor stem but lacked other conserved structural features of pre-tRNA were recognized and

cleaved, although with greater rates of miscleavage (71). Together, these data support an induced-fit mechanism for pre-tRNA cleavage, where initial association of RNase P to pre-tRNA at regions of the tRNA body induce conformational changes that facilitate proper binding and cleavage of the 5' leader.

1.4.3 Substrate Recognition by PRORP

Prior to the work described in the following chapters, little was known about the specific interactions that facilitate PRORP pre-tRNA recognition. Substrate specificity studies of *At*PRORP isoforms demonstrated little change in catalytic efficiency with a variety of substrates, indicating they shared a similar substrate recognition mechanism (72). Furthermore, these studies revealed that 5' leader and 3' trailer length were not important determinants for pre-tRNA recognition. Functional contacts with these extraneous sequences are not present, as pre-tRNA substrates with either 1-nt leaders or trailers were efficiently bound and processed.

The PPR domain, a known RNA binding domain, is essential for substrate recognition and processing. Various deletions of PPR motifs greatly reduced binding affinities and, in many cases, eliminated cleavage activity (73). RNA footprinting experiments have shown that conserved nucleotides (G18, G19, C56, and C57), that form tRNA tertiary interactions in the D-/T-loop region, are protected in the presence of *At*PRORP1 (73). These data imply that, like RNA-dependent RNase P, direct interactions with this region play a role in pre-tRNA recognition. Modeling of pre-tRNA bound to the structure of PRORP suggests that the D-/T-loop region may be recognized by the PPR motifs. In other PPR proteins, tandem PPR motifs bind single-stranded RNA (ssRNA) in a sequence-specific manner (74). This is achieved through modular binding, where amino acids 4 and 34 of each PPR motif recognize one specific nucleotide. It is unclear how this may pertain to PRORP recognition as the set of pre-tRNA sequences bound by PRORP are

not conserved. Furthermore, the D-/T-loop region predicted to interact with the PPR motifs, enters a structural fold where nucleotides are not as available to form interactions as compared to ssRNA.

As previously mentioned, many tRNAs encoded in the mitochondria of metazoans contain unique unorthodox primary structures that abolish canonical tertiary interactions. Therefore, substrate recognition strategies of metazoan mitochondrial RNase P may deviate from those of other RNase P enzymes. This perceived need for a unique substrate recognition mechanism may explain the necessity for additional protein subunits required for catalysis.

1.5 Conclusion

Collectively, the past half a century work on RNase P enzymes has produced the discovery of the first true RNA catalyst, an example of RNA nuclease activity, insights into tRNA maturation and biogenesis, and a unique opportunity to compare the catalytic strategies of RNA and protein enzymes. There are numerous reasons why further study of this essential enzyme is a worthwhile endeavor. It has been shown that changes in human mitochondrial RNase P activity may contribute to several forms of mitochondrial dysfunction (75), and bacterial RNase P may present a suitable antibiotic target. Furthermore, increased understanding of RNase P mechanisms and function may provide insight into the shift from RNA to protein-based catalysis predicted in the RNA world theory. Finally, the ubiquitous presence of RNase P and its function as a necessary component for the translation of biological information, make understanding of its molecular mechanism of function a significant contribution to our understanding life.

1.6 Objective

Previous work on RNA-based and protein-only RNase P revealed that both compositional types of the enzyme share a near identical chemical mechanism. Yet clarification of the molecular recognition mechanism of PRORP remained poorly understood. Thus, the following work entails

efforts to elucidate the substrate recognition mechanism of eukaryotic PRORP, focusing on contributions within the RNA binding PPR domain. Chapter 2 showcases a biochemical investigation of *At*PRORP1 PPR domain interactions with pre-tRNA. Site-directed mutagenesis, nucleic acid binding theory and cross-linking data indicate several specific interactions which are combined with past X-ray structures to produce a model of PRORP bound to pre-tRNA. Chapter 3 describes the crystal structure of the PRORP1 PPR domain bound to tRNA. This structure validates many of the interactions predicted by the model in Chapter 2 and reveals a new mechanism of RNA recognition for PPR proteins. Lastly, Chapter 4 discusses the significance and implications of this work and describes potential future directions of study.

1.7 References

1. Crick,F.H.C., Brenner,S., Klug,A. and Pieczenik,G. (1976) A speculation on the origin of protein synthesis. *Orig. Life*, 10.1007/BF00927934.
2. Walker,S.C. and Engelke,D.R. (2006) Ribonuclease P: The evolution of an ancient RNA enzyme. *Crit. Rev. Biochem. Mol. Biol.*, **41**, 77–102.
3. Hopper,A.K. and Huang,H.-Y. (2015) Quality Control Pathways for Nucleus-Encoded Eukaryotic tRNA Biosynthesis and Subcellular Trafficking. *Mol. Cell. Biol.*, **35**, 2052–2058.
4. Randau,L. (2012) RNA processing in the minimal organism Nanoarchaeum equitans. *Genome Biol.*, 10.1186/gb-2012-13-7-r63.
5. Randau,L., Schröder,I. and Söll,D. (2008) Life without RNase P. *Nature*, **453**, 120–123.
6. Ikeda,H. (1971) In Vitro synthesis of tRNATyr precursors and their conversion to 4S RNA. *Nat. New Biol.*, 10.1038/newbio234198a0.
7. Robertson,H.D., Webster,R.E. and Zinder,N.D. (1968) Purification and properties of ribonuclease III from Escherichia coli. *J. Biol. Chem.*, **243**, 82–91.
8. Guerrier-Takada,C., Gardiner,K., Marsh,T., Pace,N. and Altman,S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, **35**, 849–857.
9. Altman,S. (1990) Nobel lecture. Enzymatic cleavage of RNA by RNA. *Biosci. Rep.*, **10**, 317–37.
10. Hsieh,J., Koutmou,K.S., Rueda,D., Koutmos,M., Walter,N.G. and Fierke,C.A. (2010) A divalent cation stabilizes the active conformation of the B. Subtilis RNase P·Pre-tRNA complex: A role for an inner-sphere metal ion in RNase P. *J. Mol. Biol.*, **400**, 38–51.
11. Koutmou,K.S., Day-Storms,J.J. and Fierke,C.A. (2011) The RNR motif of B. subtilis RNase P protein interacts with both PRNA and pre-tRNA to stabilize an active conformer. *RNA*, **17**, 1225–1235.
12. Klemm,B.P., Karasik,A., Kaitany,K.J., Shanmuganathan,A., Henley,M.J., Thelen,A.Z., Dewar,A.J.L., Jackson,N.D., Koutmos,M. and Fierke,C.A. (2017) Molecular recognition of pre-tRNA by Arabidopsis protein-only Ribonuclease P. *RNA*, **23**.
13. Darr,S.C., Brown,J.W. and Pace,N.R. (1992) The varieties of ribonuclease P. *Trends Biochem. Sci.*, **17**, 178–82.
14. Xiao,S., Scott,F., Fierke,C.A. and Engelke,D.R. (2002) Eukaryotic ribonuclease P: a plurality of ribonucleoprotein enzymes. *Annu. Rev. Biochem.*, **71**, 165–89.
15. Chamberlain,J.R., Lee,Y., Lane,W.S. and Engelke,D.R. (1998) Purification and characterization of the nuclear RNase P holoenzyme complex reveals extensive subunit overlap with RNase MRP. *Genes Dev.*, **12**, 1678–1690.
16. Lygerou,Z., Mitchell,P., Petfalski,E., Séraphin,B. and Tollervey,D. (1994) The POP1 gene encodes a protein component common to the RNase MRP and RNase P ribonucleoproteins. *Genes Dev.*, **8**, 1423–1433.
17. Chu,S., Zengel,J.M. and Lindahl,L. (1997) A novel protein shared by RNase MRP and RNase P. *RNA*, 10.13016/m2nyzm-w3x5.
18. Rossmanith,W. and Karwan,R.M. (1998) Characterization of human mitochondrial RNase P: Novel aspects in tRNA processing. *Biochem. Biophys. Res. Commun.*, **247**, 234–241.
19. Holzmann,J., Frank,P., Löffler,E., Bennett,K.L., Gerner,C. and Rossmanith,W. (2008) RNase P without RNA: Identification and Functional Reconstitution of the Human Mitochondrial tRNA Processing Enzyme. *Cell*, **135**, 462–474.
20. Holzmann,J., Frank,P., Löffler,E., Bennett,K.L., Gerner,C. and Rossmanith,W. (2008) RNase P without RNA: Identification and Functional Reconstitution of the Human Mitochondrial tRNA

Processing Enzyme. *Cell*, **135**, 462–474.
21. Gobert,A., Gutmann,B., Taschner,A., Göringer,M., Holzmann,J., Hartmann,R.K., Rossmanith,W. and Giegé,P. (2010) A single Arabidopsis organellar protein has RNase P activity. *Nat. Struct. Mol. Biol.*, **17**, 740–744.
22. Gutmann,B., Gobert,A. and Giegé,P. (2012) PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis. *Genes Dev.*, **26**, 1022–1027.
23. Lange,A., Mills,R.E., Lange,C.J., Stewart,M., Devine,S.E. and Corbett,A.H. (2007) Classical nuclear localization signals: Definition, function, and interaction with importin α. *J. Biol. Chem.*, **282**, 5101–5105.
24. Lai,L.B., Bernal-Bayard,P., Mohannath,G., Lai,S.M., Gopalan,V. and Vioque,A. (2011) A functional RNase P protein subunit of bacterial origin in some eukaryotes. *Mol. Genet. Genomics*, **286**, 359–369.
25. Salavati,R., Panigrahi,A.K. and Stuart,K.D. (2001) Mitochondrial ribonuclease P activity of Trypanosoma brucei. *Mol. Biochem. Parasitol.*, **115**, 109–117.
26. Taschner,A., Weber,C., Buzet,A., Hartmann Roland K.,R.K., Hartig,A. and Rossmanith,W. (2012) Nuclear RNase P of Trypanosoma brucei: A Single Protein in Place of the Multicomponent RNA-Protein Complex. *Cell Rep.*, **2**, 19–25.
27. Howard,M.J., Liu,X., Lim,W.H., Klemm,B.P., Fierke,C.A., Koutmos,M. and Engelke,D.R. (2013) RNase P enzymes: divergent scaffolds for a conserved biological reaction. *RNA Biol.*, **10**, 909–914.
28. Nickel,A.I., Wäber,N.B., Gößringer,M., Lechner,M., Linne,U., Toth,U., Rossmanith,W. and Hartmann,R.K. (2017) Minimal and RNA-free RNase P in Aquifex aeolicus. *Proc. Natl. Acad. Sci. U. S. A.*, **114**, 11121–11126.
29. Schwarz,T.S., Wäber,N.B., Feyh,R., Weidenbach,K., Schmitz,R.A., Marchfelder,A. and Hartmann,R.K. (2019) Homologs of *aquifex aeolicus* protein-only RNase P are not the major RNase P activities in the archaea *haloferax volcanii* and *methanosarcina mazei*. *IUBMB Life*, **71**, iub.2122.
30. Loria,A. and Pan,T. (1996) Domain structure of the ribozyme from eubacterial ribonuclease P. *RNA*.
31. Green,C.J., Rivera-León,R. and Vold,B.S. (1996) The catalytic core of RNase P. *Nucleic Acids Res.*, 10.1093/nar/24.8.1497.
32. Harris,M.E., Nolan,J.M., Malhotra,A., Brown,J.W., Harvey,S.C. and Pace,N.R. (1994) Use of photoaffinity crosslinking and molecular modeling to analyze the global architecture of ribonuclease P RNA. *EMBO J.*, 10.1002/j.1460-2075.1994.tb06711.x.
33. Kazantsev,A. V., Krivenko,A.A., Harrington,D.J., Holbrook,S.R., Adams,P.D. and Pace,N.R. (2005) Crystal structure of a bacterial ribonuclease P RNA. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 13392–13397.
34. Schmitz,M. and Tinoco I.,J. (2000) Solution structure and metal-ion binding of the P4 element from bacterial RNase P RNA. *RNA*, **6**, 1212–1225.
35. Reiter,N.J., Osterman,A., Torres-Larios,A., Swinger,K.K., Pan,T. and Mondragón,A. (2010) Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature*, **468**, 784–791.
36. Spitzfaden,C., Nicholson,N., Jones,J.J., Guth,S., Lehr,R., Prescott,C.D., Hegg,L.A. and Eggleston,D.S. (2000) The structure of ribonuclease P protein from Staphylococcus aureus reveals a unique binding site for single-stranded RNA. *J. Mol. Biol.*, **295**, 105–115.
37. Kazantsev,A. V., Rambo,R.P., Karimpour,S., Santalucia,J., Tainer,J.A. and Pace,N.R. (2011)

Solution structure of RNase P RNA. *RNA*, 10.1261/rna.2563511.
38. Kazantsev,A. V., Krivenko,A.A., Harrington,D.J., Carter,R.J., Holbrook,S.R., Adams,P.D. and Pace,N.R. (2003) High-resolution structure of RNase P protein from Thermotoga maritima. *Proc. Natl. Acad. Sci. U. S. A.*, 10.1073/pnas.0932597100.
39. Marquez,S.M., Harris,J.K., Kelley,S.T., Brown,J.W., Dawson,S.C., Roberts,E.C. and Pace,N.R. (2005) Structural implications of novel diversity in eucaryal RNase P RNA. *RNA*, **11**, 739–751.
40. Howard,M.J., Lim,W.H., Fierke,C. a. and Koutmos,M. (2012) Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5' processing. *Proc. Natl. Acad. Sci.*, **109**, 16149–16154.
41. Karasik,A., Shanmuganathan,A., Howard,M.J., Fierke,C.A. and Koutmos,M. (2016) Nuclear Protein-Only Ribonuclease P2 Structure and Biochemical Characterization Provide Insight into the Conserved Properties of tRNA 5′ End Processing Enzymes. *J. Mol. Biol.*, **428**, 26–40.
42. Reinhard,L., Sridhara,S. and Hallberg,B.M. (2015) Structure of the nuclease subunit of human mitochondrial RNase P. *Nucleic Acids Res.*, 10.1093/nar/gkv481.
43. Li,F., Liu,X., Zhou,W., Yang,X. and Shen,Y. (2015) Auto-inhibitory Mechanism of the Human Mitochondrial RNase P Protein Complex. *Sci. Rep.*, **5**, 9878.
44. Van Laer,B., Roovers,M., Wauters,L., Kasprzak,J.M., Dyzma,M., Deyaert,E., Kumar Singh,R., Feller,A., Bujnicki,J.M., Droogmans,L., *et al.* (2015) SUP: Structural and functional insights into tRNA binding and adenosine N1-methylation by an archaeal Trm10 homologue. *Nucleic Acids Res.*, 10.1093/nar/gkv1369.
45. Oerum,S., Roovers,M., Rambo,R.P., Kopec,J., Bailey,H.J., Fitzpatrick,F., Newman,J.A., Newman,W.G., Amberger,A., Zschocke,J., *et al.* (2018) Structural insight into the human mitochondrial tRNA purine N1-methyltransferase and ribonuclease P complexes. *J. Biol. Chem.*, 10.1074/jbc.RA117.001286.
46. Liu,X., Wu,N., Shanmuganathan,A., Klemm,B.P., Howard,M.J., Lim,W.H., Koutmos,M. and Fierke,C.A. (2019) Kinetic mechanism of human mitochondrial RNase P. *bioRxiv*, 10.1101/666792.
47. Kole,R. and Altman,S. (1981) Properties of Purified Ribonuclease P from Escherichia coli. *Biochemistry*, **20**, 1902–1906.
48. Kazakov,S. and Altman,S. (1992) A trinucleotide can promote metal ion-dependent specific cleavage of RNA. *Proc. Natl. Acad. Sci. U. S. A.*, 10.1073/pnas.89.17.7939.
49. Altman,S. (1989) Ribonuclease P: an enzyme with a catalytic RNA subunit. *Adv. Enzymol. Relat. Areas Mol. Biol.*, **62**, 1–36.
50. Kurz,J.C. and Fierke,C.A. (2002) The affinity of magnesium binding sites in the Bacillus subtilis RNase P·Pre-tRNA complex is enhanced by the protein subunit. *Biochemistry*, **41**, 9545–9558.
51. Kurz,J.C., Niranjanakumari,S. and Fierke,C.A. (1998) Protein component of Bacillus subtilis RNase P specifically enhances the affinity for precursor-tRNA(Asp). *Biochemistry*, **37**, 2393–2400.
52. Koutmou,K.S., Casiano-Negroni,A., Getz,M.M., Pazicni,S., Andrews,A.J., Penner-Hahn,J.E., Al-Hashimia,H.M. and Fierke,C.A. (2010) NMR and XAS reveal an inner-sphere metal binding site in the P4 helix of the metallo-ribozyme ribonuclease P. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 2479–2484.
53. Howard,M.J., Klemm,B.P., Fierke,C. a, Arbor,A., Arbor,A. and Arbor,A. (2015) Enzymatic Mechanism of Protein-only RNase P. 10.1074/jbc.M115.644831.
54. Hernandez-Cid,A., Aguirre-Sampieri,S., Diaz-Vilchis,A. and Torres-Larios,A. (2012) Ribonucleases P/MRP and the expanding ribonucleoprotein world. *IUBMB Life*, **64**, 521–528.

55. Marvin,M.C., Clauder-Münster,S., Walker,S.C., Sarkeshik,A., Yates,J.R., Steinmetz,L.M. and Engelke,D.R. (2011) Accumulation of noncoding RNA due to an RNase P defect in Saccharomyces cerevisiae. *RNA*, 10.1261/rna.2737511.
56. Dirheimer,G., Keith,G., Dumas,P. and Westhof,E. (2014) Primary, Secondary, and Tertiary Structures of tRNAs. In *tRNA*.
57. Suzuki,T., Nagao,A. and Suzuki,T. (2011) Human Mitochondrial tRNAs: Biogenesis, Function, Structural Aspects, and Diseases. *Annu. Rev. Genet.*, **45**, 299–329.
58. Hsieh,J., Andrews,A.J. and Fierke,C.A. (2004) Roles of Protein Subunits in RNA-Protein Complexes: Lessons from Ribonuclease P. *Biopolymers*, **73**, 79–89.
59. Smith,J.K., Hsieh,J. and Fierke,C.A. (2007) Review: Importance of RNA-protein interactions in bacterial ribonuclease P structure and catalysis. *Biopolymers*, 10.1002/bip.20846.
60. Zahler,N.H., Christian,E.L. and Harris,M.E. (2003) Recognition of the 5′ leader of pre-tRNA substrates by the active site of ribonuclease P. *RNA*, 10.1261/rna.5220703.
61. Koutmou,K.S., Zahler,N.H., Kurz,J.C., Campbell,F.E., Harris,M.E. and Fierke,C.A. (2010) Protein-Precursor tRNA Contact Leads to Sequence-Specific Recognition of 5′ Leaders by Bacterial Ribonuclease P. *J. Mol. Biol.*, 10.1016/j.jmb.2009.11.039.
62. Wegscheid,B. and Hartmann,R.K. (2006) The precursor tRNA 3′-CCA interaction with Escherichia coli RNase P RNA is essential for catalysis by RNase P in vivo. *RNA*, 10.1261/rna.188306.
63. Reilly,R.M. and RajBhandary,U.L. (1986) A single mutation in loop IV of Escherichia coli SuIII tRNA blocks processing at both 5'- and 3'-ends of the precursor tRNA. *J. Biol. Chem.*
64. Kirsebom,L.A. and Altman,S. (1989) Reaction in vitro of some mutants of RNase P with wild-type and temperature-sensitive substrates. *J. Mol. Biol.*, 10.1016/0022-2836(89)90250-7.
65. Kahle,D., Wehmeyer,U. and Krupp,G. (1990) Substrate recognition by RNase P and by the catalytic M1 RNA: identification of possible contact points in pre-tRNAs. *EMBO J.*, 10.1002/j.1460-2075.1990.tb08320.x.
66. Nolan,J.M., Burke,D.H. and Pace,N.R. (1993) Circularly permuted tRNAs as specific photoaffinity probes of ribonuclease P RNA structure. *Science (80-.).*, 10.1126/science.7688143.
67. Loria,A. and Pan,T. (1997) Recognition of the T stem-loop of a pre-tRNA substrate by the ribozyme from Bacillus subtilis ribonuclease Pp. *Biochemistry*, 10.1021/bi970115o.
68. Klemm,B.P., Wu,N., Chen,Y., Liu,X., Kaitany,K.J., Howard,M.J. and Fierke,C.A. (2016) The diversity of ribonuclease P: Protein and RNA catalysts with analogous biological functions. *Biomolecules*, **6**.
69. Loria,A., Niranjanakumari,S., Fierke,C.A. and Pan,T. (1998) Recognition of a pre-tRNA substrate by the Bacillus subtilis RNase P holoenzyme. *Biochemistry*, 10.1021/bi9816507.
70. Loria,A. and Pan,T. (1999) The cleavage step of ribonuclease P catalysis is determined by ribozyme- substrate interactions both distal and proximal to the cleavage site. *Biochemistry*, 10.1021/bi990691f.
71. Brännvall,M., Kikovska,E., Wu,S. and Kirsebom,L.A. (2007) Evidence for Induced Fit in Bacterial RNase P RNA-mediated Cleavage. *J. Mol. Biol.*, 10.1016/j.jmb.2007.07.030.
72. Howard,M.J., Karasik,A., Klemm,B.P., Mei,C., Shanmuganathan,A., Fierke,C.A. and Koutmos,M. (2016) Differential substrate recognition by isozymes of plant protein-only Ribonuclease P. *RNA*, 10.1261/rna.055541.115.
73. Imai,T., Nakamura,T., Maeda,T., Nakayama,K., Gao,X., Nakashima,T., Kakuta,Y. and Kimura,M. (2014) Pentatricopeptide repeat motifs in the processing enzyme PRORP1 in Arabidopsis thaliana play a crucial role in recognition of nucleotide bases at TψC loop in

precursor tRNAs. *Biochem. Biophys. Res. Commun.*, **450**, 1541–1546.
74. Manna,S. (2015) An overview of pentatricopeptide repeat proteins and their applications. *Biochimie*, 10.1016/j.biochi.2015.04.004.
75. Yarham,J.W., Elson,J.L., Blakely,E.L., Mcfarland,R. and Taylor,R.W. (2010) Mitochondrial tRNA mutations and disease. *Wiley Interdiscip. Rev. RNA*, 10.1002/wrna.27.

Chapter 2 Molecular Recognition of Pre-tRNA by Arabidopsis Protein-only Ribonuclease P[1]

2.1 Abstract

Protein-only ribonuclease P (PRORP) is an enzyme responsible for catalyzing the 5' end maturation of precursor transfer ribonucleic acids (pre-tRNAs) encoded by various cellular compartments in many eukaryotes. PRORPs from plants act as single-subunit enzymes and have been used as a model system for analyzing the function of the metazoan PRORP nuclease subunit, which requires two additional proteins for efficient catalysis. There are currently few molecular details known about the PRORP-pre-tRNA complex. Here, we characterize the determinants of substrate recognition by the single subunit *Arabidopsis thaliana* PRORP1 and PRORP2 using kinetic and thermodynamic experiments. The salt dependence of binding affinity suggests 4–5 contacts with backbone phosphodiester bonds on substrates, including a single phosphodiester contact with the pre-tRNA 5' leader, consistent with prior reports of short leader requirements. PRORPs contain an N-terminal pentatricopeptide repeat (PPR) domain, truncation of which results

[1] The following chapter has been published in RNA: Klemm BP, Karasik A, Kaitany KJ, et al. Molecular recognition of pre-tRNA by *Arabidopsis* protein-only Ribonuclease P. *RNA*. 2017;23 (12):1860-1873. My contributions to this work include the molecular cloning and purification of several *Arabidopsis Thaliana* PRORP1 variants. I characterized these, and other variants, through performing fluorescence anisotropy binding assays and single-turnover kinetic assays at varying salt concentrations for analysis of RNA phosphodiester backbone – protein binding interactions. Additionally, I produced fluorescently labeled pre-tRNA used within this chapter (*B. subtilis* pre-tRNA^{ASP}) through *in vitro* transcription. I took part in data analysis, writing and editing of the final manuscript and was responsible for running additional assays that were requested by reviewers during the submission process.

in > 30-fold decrease in substrate affinity. While most PPR-containing proteins have been implicated in single-stranded sequence specific RNA recognition, we find that the PPR motifs of PRORPs recognize pre-tRNA substrates differently. Notably, the PPR domain residues most important for substrate binding in PRORPs do not correspond to positions involved in base recognition in other PPR proteins. Several of these residues are highly conserved in PRORPs from algae, plants, and metazoans, suggesting a conserved strategy for substrate recognition by the PRORP PPR domain. Furthermore, there is no evidence for sequence specific interactions. This work clarifies molecular determinants of PRORP-substrate recognition and provides a new predictive model for the PRORP-substrate complex.

2.2 Introduction

Ribonuclease P (RNase P) enzymes are essential endonucleases with diverse macromolecular composition that are responsible for catalyzing the maturation of the 5' end of pre-tRNA (Howard *et al.* 2013). In many biological settings, RNase P is a ribonucleoprotein complex containing a large catalytic RNA capable of processing pre-tRNAs *in vitro* (Guerrier-Takada *et al.* 1983). Additionally, one or more associated protein components are required for function *in vivo* (Walker and Engelke 2006; Marvin and Engelke 2009). In bacterial RNase P, the protein subunits increase substrate affinity and the ability of divalent metal ions to bind at specific sites (Crary *et al.* 1998; Kurz *et al.* 1998; Niranjanakumari *et al.* 1998; Kurz and Fierke 2002).

In many eukaryotic species, including protists, algae, land plants, and metazoans, protein-only RNase Ps (PRORPs) have been identified (Holzmann *et al.* 2008; Gobert *et al.* 2010; Lai *et al.* 2011; Taschner *et al.* 2012). Human mitochondrial RNase P (mtRNase P) was the first PRORP described and it requires 2 additional protein subunits for activity (Holzmann *et al.* 2008). These subunits are an m^1G/A_9 tRNA-methyltransferase (TRMT10C, also MRPP1) and a hydroxysteroid

dehydrogenase/reductase (HSD17B10, also MRPP2), which form a sub-complex (Holzmann *et al.* 2008; Vilardo *et al.* 2012). MRPP1 and MRPP2 are proposed to contribute primarily to substrate recognition.

In contrast to the metazoan PRORP, the PRORPs from algae, protists, and plants do not require additional subunits for efficient catalysis *in vitro* (Gobert *et al.* 2010; Lai *et al.* 2011; Gutmann *et al.* 2012; Taschner *et al.* 2012; Sugita *et al.* 2014; Howard *et al.* 2015; Bonnard *et al.* 2016), suggesting differences in substrate recognition. The three PRORPs from *A. thaliana* are designated PRORP1–3. *At*PRORP1 localizes to the mitochondria and chloroplasts where it is responsible for catalyzing pre-tRNA maturation (Gobert *et al.* 2010), while *At*PRORP2 and *At*PRORP3 co-localize to the nucleus and are not fully redundant in nuclear pre-tRNA processing (Gutmann *et al.* 2012). *At*PRORP1 utilizes a metal ion-dependent mechanism similar to the mechanism of the ribozyme, relying on ionization of metal-bound waters for nucleophile activation in catalysis (Chen *et al.* 1997; Howard *et al.* 2015). Given the additional mechanistic information and the relative simplicity of the *At*PRORPs, they have been used as a model system to study PRORP-substrate molecular recognition.

PRORPs contain a unique three-domain architecture (Figure 2-1A). An N-terminal pentatricopeptide repeat (PPR) domain significantly enhances the affinity for substrate and truncation of the first 3–4 repeats abolishes catalytic activity (Howard *et al.* 2012; Imai *et al.* 2014; Karasik *et al.* 2016). Thus, the PPR domain is proposed to both bind and orient substrate with respect to the metallonuclease domain (Howard *et al.* 2012). In addition, the nuclease domain is a member of the Nedd4-BP1, YacP nuclease (NYN) family (Anantharaman and Aravind 2006). Lastly, a bipartite CC/HC Zn^{2+}-binding domain flanks the NYN domain (Howard *et al.* 2012). Our

current understanding of how each domain, in particular the PPR domain (Figure 2-1B, Figure 2-8), contributes to PRORP substrate recognition is limited.

Previous results suggest differences in substrate recognition between the bacterial RNA-dependent RNase P and PRORP. Unlike the ribozyme, the PRORP active site metal ions apparently do not contact the pro-R_P oxygen (Pavlova *et al.* 2012), but rather contact the pro-S_P oxygen of the scissile phosphodiester bond (Walczyk *et al.* 2016). Furthermore, while the 3'-CCA is specifically recognized by bacterial RNase P RNA, it is either inhibitory or immaterial to *At*PRORP activity (Gobert *et al.* 2013; Brillante *et al.* 2016; Mao *et al.* 2016). Additionally, *At*PRORP1 and *At*PRORP3 do not significantly contact either the 5' leader sequence beyond N_{-2} or the 3' trailer (Figure 2-2A); these regions do not alter substrate affinity or catalytic activity (Brillante *et al.* 2016; Howard *et al.* 2016). The minimal 5' and 3' end interactions indicate that PRORP substrate recognition lies primarily within the tRNA body.

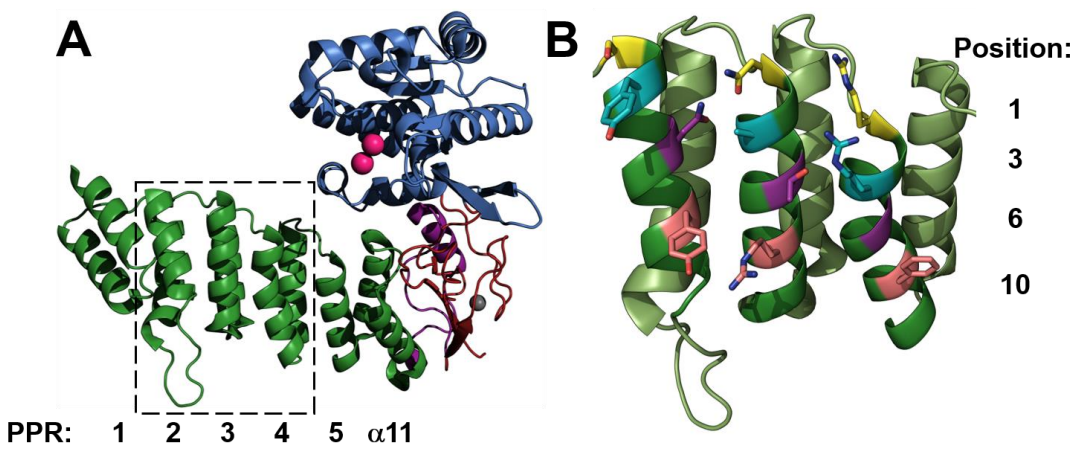


Figure 2-1 Structure of *Arabidopsis thaliana* PRORP1 (PDB 4g24), generated in PyMOL (Schrödinger 2010).

(**A**) Overall architecture of *At*PRORP1. PPR motifs are in green and numbered from the N-terminal end, a plant-specific helical insertion in purple (α12–13), central domain in red, and NYN domain in blue. Mn^{2+} ions displayed as pink spheres and Zn^{2+} ion displayed as a gray sphere. The region marked by the dashed box is expanded in panel B. (**B**) *At*PRORP1 PPR domain motifs 2–4,

generated in PyMOL. For each PPR motif, position 1 is colored yellow, position 3 is colored cyan, position 6 is colored purple, and position 10 is colored pink.

Previously, a nuclease footprinting assay demonstrated that there was significant protection of bases in the D- and TψC-loops (Figure 2-2A) by *At*PRORP1 (Gobert *et al.* 2013). Given these data and the likelihood that the NYN domain binds at the scissile phosphodiester bond, it was proposed that the PPRs recognized the pre-tRNA elbow, the structure formed by interaction between the D- and TψC-loops (Gobert *et al.* 2013). However, this proposal remains to be tested. Recent attempts to alter base specificity of the PPR domain in *At*PRORP3 were unsuccessful (Brillante *et al.* 2016). Furthermore, while the TψC-arm is sufficient for recognition and catalysis by plant PRORPs, the presence of a D-arm increases the affinity significantly (Brillante *et al.* 2016; Howard *et al.* 2016). These data provide a basis for examining the features of PRORP that contribute to recognition of pre-tRNA.

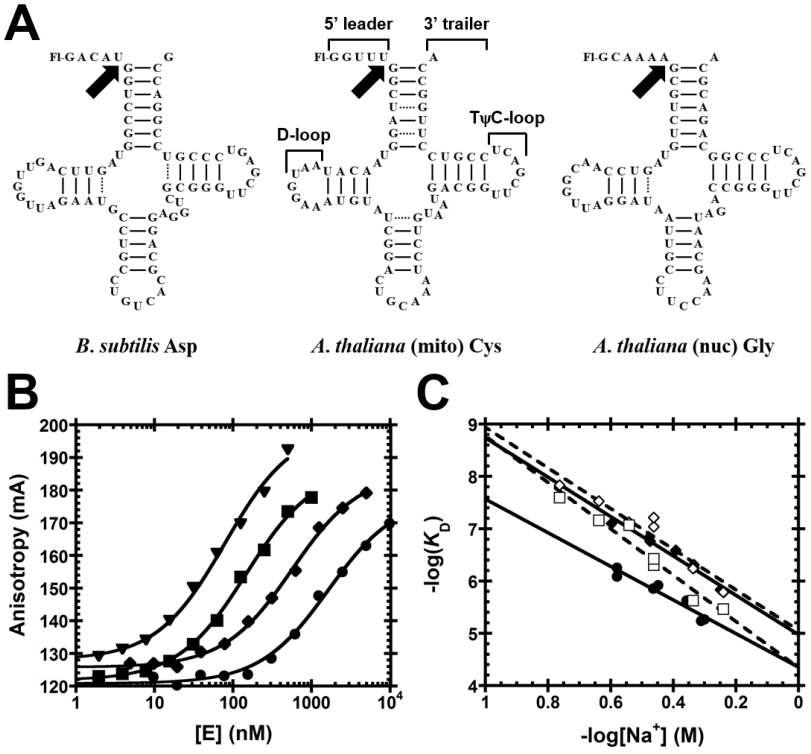


Figure 2-2 Substrates used for functional assays.

(**A**) Substrates containing a 5'-fluorescein label include pre-tRNA^{Asp} from Bacillus subtilis and Arabidopsis thaliana pre-tRNA^{Cys} and pre-tRNA^{Gly} from the mitochondrial and nuclear genomes, respectively. Structural features of pre-tRNA are detailed on pre-tRNA^{Cys}, including the 5' leader, 3' trailer, and the D- and TψC-loops. Black arrows indicate canonical RNase P cleavage site. (**B**) Fluorescence anisotropy binding curves for AtPRORP1 binding to B. subtilis pre-tRNA^{Asp} (log scale). A hyperbola (Equation 1, Materials and Methods) was fit to the data. Data were measured in 30 mM MOPS pH 7.8, 1 mM TCEP, and 20 mM $CaCl_2$ with 250 mM (▼), 330 mM (■), 450 mM (♦), and 550 mM NaCl (●). (**C**) Na^+-dependence of *At*PRORP affinity. Equation 4 (Materials and Methods) was fit to the data. The slope of the line (Z) reports on the apparent number of ionic interactions with substrate phosphodiester bonds, while the intercept [$\log(K_0)$] reports on the non-ionic contributions to affinity. Data include AtPRORP1 binding to pre-tRNA^{Asp} (♦) and pre-tRNA^{Cys} (●) in 20 mM Ca^{2+}, as well as *At*PRORP2 in 6 mM Ca^{2+} binding to pre-tRNA^{Asp} (◊) and pre-tRNA^{Gly} (□).

A model of substrate-bound *At*PRORP1 was previously generated using molecular dynamics and includes the PPR domain docked to the TψC-loop (Imai *et al.* 2014). The authors assumed that PRORPs use the recognition strategy employed by several single-stranded RNA binding PPR proteins. The ssRNA-binding PPR proteins recognize nucleobases utilizing residues in two tandem repeats at positions 6 and 1' (Figure 2-1B, Figure 2-8), as well as hydrophobic amino acids at position 3 that contribute to binding affinity by van der Waals or stacking interactions (Barkan *et al.* 2012; Yagi *et al.* 2013). Cleavage assays catalyzed by *At*PRORP1 indicated that mutations to position 6 of PPR motifs 2, 3, and 4 reduce activity modestly (≤ 70% reduction) (Imai *et al.* 2014). However, the full suite of PPR residues important for PRORP substrate recognition remains to be identified.

Here, we characterize the mode of substrate binding and recognition by the highly conserved *At*PRORP1 and 2 using a variety of biochemical techniques. The salt-dependence of pre-tRNA affinity indicate that *At*PRORP1 and *At*PRORP2 make at least four direct contacts to substrate backbone phosphodiester bonds, including a single phosphodiester bond contact with the pre-tRNA leader. Importantly, these interactions with the backbone are not sequence specific. However, the salt-dependence of affinity for mature tRNA also demonstrates that a significant portion of the affinity for substrate stems from interactions with the sugars and/or bases in the body

of the substrate, in a contrast to the bacterial ribozyme, which makes far more contacts to the 5' leader and 3' CCA. To test whether *At*PRORP1 uses canonical PPR-nucleobase interactions, we mutated both residues in the PRORP PPR domains and nucleotides in a pre-tRNA substrate and assessed how the mutations impact the PRORP-pre-tRNA affinity. In contrast to other known PPR proteins, PRORP does not exhibit demonstrable sequence selectivity for substrate affinity, suggesting that substrate recognition relies instead on the three-dimensional structure of pre-tRNA. These experiments provide a biochemical framework for understanding molecular recognition of complex RNA structures by the non-canonical PPRs of plant PRORPs.

2.3 Materials and Methods

Enzyme preparation

Variants of Δ76 *At*PRORP1 and full length *At*PRORP2 were generated by site directed mutagenesis (Hutchison *et al.* 1978). Sequences were verified at the University of Michigan DNA Sequencing Core facility. Variants were expressed in Rosetta™, Rosetta 2™ or BL21(DE3) *E. coli* (Novagen/EMD Millipore) from the T7 promoter on a pETM-11 (encoding His_6-TEV-*At*PRORP1) or pMCSG7 (His_6-TEV-*At*PRORP2) vector in LB media with 50 μg/mL kanamycin and 33 μg/mL chloramphenicol for selection of pETM-11 and pRARE (a plasmid encoding rare-codon tRNAs in the Rosetta cell lines) or 100 μg/mL ampicillin for selection of pMCSG7. Wild type Δ76 *At*PRORP1 and full length *At*PRORP2 and variants of these enzymes were purified as described previously (Howard *et al.* 2012; Howard *et al.* 2015; Karasik *et al.* 2016).

Substrate preparation

Substrates were prepared as described previously (Howard *et al.* 2012; Howard *et al.* 2015). Briefly, substrates were synthesized by run-off transcription from restriction-digested plasmid encoding pre-tRNA, a PCR-amplified template DNA, or a commercially synthetized ultramer oligo (IDT) (Milligan and Uhlenbeck 1989). *In vitro* transcription was carried out in the presence of 5'-O-monophosphorothioate guanosine (GMPS) in 5:1 excess of GTP. The pre-tRNA containing a 5'-GMPS was reacted with 5-iodoacetamidofluorescein (5-IAF) to generate a 5'-fluorescein label. The pre-tRNA product was gel purified using 12% urea-PAGE and substrate was eluted from the gel using the crush-soak method (Milligan and Uhlenbeck 1989). The purified pre-tRNAs were washed and concentrated using 10 kDa MWCO Amicon® Ultra Centrifugal Filters, then ethanol precipitated. Substrate stocks were resuspended in 10 mM tris(hydroxymethyl)aminomethane (Tris) pH 8.0 with 1 mM EDTA, quantified by absorbance and stored at –20 or –80°C. The extinction coefficients at 260 nm for total RNA concentration are: 685000 M^{-1} cm^{-1} for *Bacillus subtilis* pre-$tRNA^{Asp}$ (experimentally determined by alkaline hydrolysis), 674390 M^{-1} cm^{-1} for *A. thaliana* mitochondrial pre-$tRNA^{Cys}$ (experimentally determined by alkaline hydrolysis), and 870700 M^{-1} cm^{-1} for *A. thaliana* nuclear pre-$tRNA^{Gly}$ (calculated). The fluorescein concentration was measured at 492 nm (extinction coefficient = 78000 M^{-1} cm^{-1}). Variants of pre-$tRNA^{Asp}$ were generated by site-directed mutagenesis (Hutchison *et al.* 1978). Sequences were verified at the University of Michigan DNA Sequencing Core facility. Immediately before initiating an assay, substrates were thawed, diluted with H_2O, and heated at 95°C for 60–90 seconds. Substrates were re-folded by cooling to 25°C for ≥ 10 min, then incubating with buffer (as specified for each assay) for ≥ 10 min.

Anisotropy binding assays

Thermodynamic binding assays were performed in a 96-well plate format as previously described (Howard *et al.* 2012; Howard *et al.* 2015). Briefly, WT *At*PRORP1 was serially diluted and mixed 1:1 with a concentration of pre-folded pre-tRNA containing a 5-fluorescein. In all experiments the maximum enzyme concentration ([P]) was at least three times greater than the K_D and the pre-tRNA concentration at least five times lower than the K_D. In all cases, the data were well-described by a hyperbolic binding curve (Equation 1). Reactions were incubated at 28 ± 1°C in 30 mM MOPS pH 7.8, and 1 mM TCEP (*At*PRORP1) or 1 mM DTT (*At*PRORP2). The NaCl concentration was varied between 0.025–1.0 M. Unless otherwise specified, assays contained 20 or 6 mM $CaCl_2$ for *At*PRORP1 and *At*PRORP2, respectively. The $CaCl_2$ concentrations were chosen to maintain the K_D values in the measurable range. For *At*PRORP1, decreasing the $CaCl_2$ concentration to 6 mM increased the K_D values (65%) but altered the slope of the Na^+-dependence < 10% (data not shown). When varying the Na^+ salt, we maintained constant Na^+ at 330 mM. Changes in anisotropy of the 5'-fluorescein-pre-tRNA was measured with a Tecan Ultra plate reader with polarizing filters using excitation and emission wavelengths of 485 and 535 nm, respectively. Readings were taken 3–5 times over the course of 15–20 minutes to ensure that the reading was stable.

$$FA = FA_0 + \frac{\Delta FA \bullet [P]}{[P] + K_D} \qquad (1)$$

Single-turnover assays

Single-turnover kinetic assays for *At*PRORP1 were performed in a stopped-format as previously described (Howard *et al.* 2012; Howard *et al.* 2015). Briefly, enzyme was mixed with pre-folded *B. subtilis* pre-$tRNA^{Asp}$ with a 5-nt leader and a 5'-fluorescein to final concentrations of 5 μM and 30 nM, respectively. For R184A assays, 30 nM substrate was incubated with 1–50

μM enzyme. Reactions were incubated at 25°C in 30 mM MOPS pH 7.8, 1 mM TCEP, with $MgCl_2$ and NaCl varied as indicated for a given assay. Aliquots were removed at various times and mixed 1:1 with a 2x quench dye (6 M urea (MP Biochemicals), 100 mM EDTA (Acros Organics), 0.1% bromophenol blue (BPB; Fisher Scientific), 0.1% xylene cyanol (XC; United States Biochemical Corporation), and 2 μg/μL bulk yeast tRNA (Fisher Scientific)). Products were resolved from substrate by fractionation on a ≥ 20% urea-PAGE and the gels were scanned using a Typhoon 9410 (GE Life Sciences) in fluorescence mode with a 532 nm green laser and fluorescein emission filter. Assays for *At*PRORP2 were carried out using the same conditions, but changes in polarization upon cleavage were detected by ClarioStar (BMG Labtech) in 96-well plate format. The observed rate constants (k_{obs}) were determined by quantifying the fraction product using ImageQuant 5.2 software and fitting a single exponential (Equation 2, where *A* is the endpoint, *B* is the amplitude, and *t* is the time) to the data using KaleidaGraph 4.0 software. At low concentrations of NaCl (below ≈ 90 mM), the 5' leader product degraded after it appeared and did not accumulate to 100%. A double exponential was fit to these data and the k_{obs} from the phase with increasing product is reported. The IC_{50} for inhibition of *At*PRORP1 by NaCl was determined by fitting Equation 3 to the dependence of the STO k_{obs} on the NaCl concentration (as described in the Results section).

$$\text{Fraction Product} = A - B(e^{-k_{obs}*t}) \qquad (2)$$

$$k_{obs} = \frac{k_{max}}{\left(1+\left(\frac{[Na^+]}{IC_{50}}\right)^n\right)} \qquad (3)$$

Sodium dependence

Equation 4 is an approximation of Equation S1 (shown in the supporting methods) that describes the dependence of the affinity on cations when effects from pH, anions, and divalent ions are negligible or can otherwise be precluded by maintaining constant pH and divalent ions

and observing the log-linear region of the decreasing affinity. The dependent variable is the monovalent cation concentration ($[M^+]$). The parameters include a "standard affinity" at 1 M M^+ (K_0), the apparent number of phosphodiester bonds on the substrate interacting with the protein (Z), and the fraction of phosphodiester bonds in the nucleic acid that thermodynamically associate with a monovalent ion (φ). When divalent cations are varied in the absence of monovalent ions, the slope is distinguished by replacing φ with ϕ. Standard affinity values were converted to energetic values using Gibbs free energy definitions and assuming equilibrium conditions (Equation 5), for which R is the gas constant (1.987 cal K^{-1} mol^{-1}) and T is the temperature (300K for our assays).

$$-\log K_D = \log K_0 - Z\varphi \bullet \log[M^+] \quad (4)$$

$$\Delta G = -RT \ln K_0 \quad (5)$$

Model building

Crystal structures of *At*PRORP1 (PDB ID: 4g23) and *At*PRORP2 (PDB ID: 5diz) and yeast $tRNA^{Phe}$ (PDB ID: 1ehz) were used to model the elbow region of pre-tRNA bound to the proteins. Initial models were obtained using ZDOCK server (Pierce *et al.* 2014) and these were processed through iterative rounds of manual adjustment by PyMOL (Schrödinger 2010). The model amino acid or nucleotide geometry regularization and use of allowed side chain rotomers were corrected with Coot (Emsley *et al.* 2010). The coordinates of these models are available upon request.

Reagents

Reagents used in this manuscript include 3-morpholinopropane-1-sulfonic acid (MOPS; Acros Organics), tris(2-carboxyethyl)phosphine (TCEP; Gold Biotechnology, Inc.), 1,4-dithio-D-threitol (DTT; Gold Biotechnology, Inc.) 5-iodoacetamidofluorescein (5-IAF; Life Technologies), ethylenediaminetetraacetic acid (EDTA; Acros Organics), tris(hydroxymethyl)aminomethane

(Tris; Fisher Scientific), magnesium chloride ($MgCl_2$; Fisher Scientific), calcium chloride ($CaCl_2$; Sigma-Aldrich), sodium chloride (NaCl; Sigma-Aldrich), sodium acetate (NaAc; Fisher Scientific), sodium bromide (NaBr; Sigma-Aldrich), sodium iodide (NaI; Sigma-Aldrich), sodium nitrate ($NaNO_3$; Sigma-Aldrich), sodium sulfate (Na_2SO_4; Fisher Scientific), 5'-O-monophosphorothioate guanosine (GMPS, synthesized), nucleoside triphosphates (NTPs; Sigma-Aldrich), kanamycin (Acros Organics), chloramphenicol (Acros Organics), urea (MP Biochemicals), bromophenol blue (BPB; Fisher Scientific), xylene cyanol (XC; United States Biochemical Corporation), SYPRO® Orange Protein Gel Stain (Invitrogen), bulk yeast tRNA (Fisher Scientific), and the SequaGel UreaGel system (National Diagnostics).

Protein-nucleic acid interactions

Equation 6 was adapted from Equation 18 of deHaseth, *et al*. 1977 and describes the effects of solution ions on the binding affinity between proteins and nucleic acids (deHaseth *et al.* 1977). Variables potentially affecting the measured affinity include pH ($K_H[H^+]$), anion concentration ($K_X[X^-]$), monovalent cation concentration ($[M^+]$), and divalent cation concentration ($[S]/[S_0]$). Parameters include a "standard affinity" under 1 M M^+ (K_0), the number of proton donating groups on the protein directly interacting with substrate phosphates (r), the number of specifically-occupied anion binding sites on the protein (a), the number of phosphates on the substrate directly interacting with the protein (Z), the fraction of phosphates in the nucleic acid that thermodynamically associate with a monovalent ion (φ), and the competition of M^+ with M^{2+} for occupancy on backbone phosphates given by the total concentration of nucleotides [S] and the concentration of nucleotides that are bound as they would with no M^{2+} in solution [S_0]. The term $[S]/[S_0]$ is described by Equation 7.

$$-log\,K_{\mathrm{D}} = log\,K_0 - r \bullet log\left(\frac{1+K_H[H^+]}{K_H[H^+]}\right) - a \bullet log(1 + K_X[X^-]) - Z\varphi \bullet log[M^+] - Z \bullet log\left(\frac{[S]}{[S_0]}\right)$$

(6)

$$\frac{[S]}{[S_0]} = \frac{1}{2}\left(1 + \sqrt{1 + 4K_A^M[M^{2+}]}\right) \quad (7)$$

Thermofluor assays

Melting curves for each variant at 5 μM enzyme were measured in 30 mM MOPS, pH 7.8, 150 mM NaCl, and 1 mM TCEP, with 20X SYPRO® Orange. Fluorescence intensity (*I*) was monitored using λ_{Ex} = 485 nm and λ_{Em} = 625 nm. The melting temperature (T_m) was determined by fitting equation 8 was fit to the data to the maximum intensity, as previously described (Ericsson *et al.* 2006), for which *A* is the initial intensity, *B* is the maximum intensity, *T* is the temperature in Celsius, and *C* is a slope factor.

$$I = \left(A + \frac{(B-A)}{\left(1+e^{(Tm-T)/C}\right)}\right) \quad (8)$$

Circular dichroism assays.

Far UV-CD spectra are presented as the average of 5 accumulations for 10 μM enzyme measured at 25°C in 10 mM phosphate buffer, pH 7.2. The CD measurements were carried out on a Jasco J-815 spectropolarimeter equipped with a circulating water bath and peltier temperature controller. Spectra were converted to mean molar residue ellipticity (θ) using equation 9 as previously described (Jayasimha *et al.* 2012), for which M_r is the mean molar mass per residue, *l* is the path-length, and *c* is the protein concentration. Overall, the characteristic features of the far UV-CD of WT *At*PRORP2 were retained in Y74S, suggesting that there are no major conformational changes when compared to WT.

$$[\theta] = \frac{\theta \bullet M_r}{10 \bullet l \bullet c} \quad (9)$$

2.4 Results

*At*PRORP-substrate recognition mode

To begin characterizing how *At*PRORPs recognize their cognate substrates, we set out to determine the general mode of substrate binding. We first measured the dependence of the protein-nucleic acid interaction on the concentration and identity of ions in solution. These data parse the dependence of affinity on ionic interactions with backbone phosphodiester bonds, compared to that of non-ionic interactions. Monovalent and divalent cations directly interact with backbone phosphodiester bonds on nucleic acids. These ions must be released for a protein to directly contact those sites, thus affinity depends on the cation concentration (Record *et al.* 1978; Barkley *et al.* 1981). Cations associate with nucleic acids through an ionic atmosphere also containing anions, which inhibits protein-nucleic acid interactions through a related screening mechanism (Record *et al.* 1978).

We measured the dependence of the substrate binding affinity of *At*PRORP1 and *At*PRORP2 on ions in solution to estimate the number of backbone phosphodiester bond contacts. We determined dissociation constants (K_D) for *At*PRORP1 and 2 by fluorescence anisotropy (FA) assays using three substrates: a *B. subtilis* pre-tRNA^{Asp} (*At*PRORP1 and 2), an *A. thaliana* mitochondrial pre-tRNA^{Cys} (*At*PRORP1), each with 5-nt leaders, and an *A. thaliana* nuclear pre-tRNA^{Gly} (*At*PRORP2) with a 6-nt leader; all substrates have a fluorescein label at the 5' end (Figure 2-2A). The pre-tRNA^{Asp} substrate has been used extensively with the bacterial ribozyme, allowing us to make direct comparisons to PRORPs, while the pre-tRNA^{Cys} is a cognate substrate for *At*PRORP1 and the pre-tRNA^{Gly} is a cognate substrate for *At*PRORP2. We obtained thermodynamic affinities (K_D) by fitting a hyperbola (Equation 1, Materials and Methods) to the data (Figure 2-2B).

The Na^+-dependence of K_D shows a linear dependence in a log-log plot (Figure 2-2C), as described by Equation 6, which was adapted from Equation 18 of deHaseth, *et al*. (deHaseth *et al.* 1977). Divalent cations are required to fold the pre-tRNA, so $CaCl_2$, which does not activate *At*PRORP1 or *At*PRORP2 (Howard *et al.* 2012; Karasik *et al.* 2016), was supplied at a constant value for each measurement. The primary effect of Ca^{2+} in our assays is to increase the anisotropy of free pre-tRNA (see the next section), while the affinity we measure at different $CaCl_2$ concentrations varied ≤ 65% at a given NaCl concentration. Thus, we selected $CaCl_2$ concentrations that allowed us to best measure the affinity under high concentrations of NaCl (20 or 6 mM $CaCl_2$ for *At*PRORP1 or *At*PRORP2, respectively). We observe minimal competition between Ca^{2+} and Na^+ for the RNA substrate under the concentrations used for the binding assays for *At*PRORP1, as evidenced by the relatively linear Na^+-dependence in the log-log plot (Figure 2-2C). We maintained constant pH during the experiments and anion effects are precluded based on the $CaCl_2$-alone and Na_2SO_4 data described below. In the absence of these effects, Equation 6 can be reduced to Equation 4 (Materials and Methods), which was fit to the data.

The slope of a $-\log(K_D)$ versus $-\log[Na^+]$ plot is given by $Z\varphi$ (Equation 4), represented as $Z\phi$ for divalent cations, where φ^{Na} is the fraction of Na^+ associated thermodynamically with each backbone phosphodiester bond and Z is the number of cations (M^+) that are released from the nucleic acid upon binding to the protein, which approximates the number of protein-phosphodiester bond contacts. Previous data suggest that the value of φ^{Na} for dsRNA and structured RNAs, such as pre-tRNA, are comparable to the φ^{Na} for dsDNA (Latt and Sober 1967; Day-Storms *et al.* 2004). Thus, we used the value for dsDNA, $\varphi^{Na} = 0.88$, in fitting Equation 4 to the data (Table 2-1). The Z values for *At*PRORP1 suggest the formation of four protein-phosphodiester bond contacts upon binding pre-tRNA. The values for *At*PRORP2 are higher,

possibly suggesting contacts with five phosphodiester groups. Furthermore, these *Z* values are fairly robust, even if we vary the approximated φ^{Na} for folded pre-tRNA. Specifically, the *Z* value for *At*PRORP2 binding to pre-tRNA^{Gly} increases to 6 if $\varphi^{Na} < 0.85$. However, *Z* will remain ≤ 5 for the other PRORP/pre-tRNA pairs until $\varphi^{Na} < 0.75$ and they are ≤ 6 until $\varphi^{Na} < 0.6$, which is likely to be well below the actual value.

Table 2-1 Na^+-dependence of binding affinity.

Enzyme	Pre-tRNA Substrate	Leader	K_D (nM)[a]	Z[b]	–log(K_0)[b]	ΔG_0 (kcal/mol)[c]
*At*PRORP1	Asp	5-nt	155 ± 20	4.3 ± 0.3	5.0 ± 0.1	–6.9 ± 0.1
		1-nt	600 ± 50	3.7 ± 0.2	4.66 ± 0.06	–6.4 ± 0.1
		0-nt	25400 ± 6100	2.8 ± 0.5	3.5 ± 0.2	–4.8 ± 0.3
	Cys	5-nt	1330 ± 120	3.7 ± 0.5	4.4 ± 0.2	–6.0 ± 0.3
*At*PRORP2	Asp	5-nt	80 ± 9	4.4 ± 0.6	5.0 ± 0.3	–6.9 ± 0.4
		0-nt	17200 ± 1100	3.2 ± 0.3	3.5 ± 0.2	–4.8 ± 0.3
	Gly	6-nt	470 ± 50	5.0 ± 0.5	4.3 ± 0.2	–5.9 ± 0.3

a: Value and error reported are from fitting a hyperbola to the results of two independent experiments in 330 mM NaCl plotted together.
b: Value and error from fitting Equation 4 to the data from Figures 2-2C and 2-4 using φ^{Na}= 0.88, as described in the Materials and Methods.
c: Calculated using $\Delta G_0 = -RT \times \ln K_0$.

The estimated substrate affinity at 1 M NaCl has been used to estimate the contribution of non-ionic interactions to affinity in model systems, using normal Gibbs free energy definitions (Equation 5) (Record *et al.* 1976). For *At*PRORP1 at 27°C, the log(K_D) at 1 M NaCl indicates values of –6.9 ± 0.1 and –6.0 ± 0.3 kcal/mol for pre-tRNA^{Asp} and pre-tRNA^{Cys}, respectively (Figure

2-2C). For *At*PRORP2, the values are –6.9 ± 0.2 and –5.9 ± 0.4 kcal/mol for pre-tRNA^{Asp} and pre-tRNA^{Gly}, respectively (Figure 2-2C). The reduced affinity for pre-tRNA^{Cys} or pre-tRNA^{Gly} compared to pre-tRNA^{Asp} represents only a minor loss of non-ionic interactions.

_At_PRORP1 does not have specific anion binding sites that compete with substrate binding

Specific anion binding sites on proteins can decrease the binding affinity of nucleic acids, in an ion-dependent manner similar to the effect of cations binding to nucleic acids. We screened the *At*PRORP1 affinity for pre-tRNA^{Asp} in several Na^+ salts at a single Na^+ concentration, but the data largely follow the lyotropic series (Table 2-2). This trend suggests that the primary effect of the anions is increasing or decreasing protein stability, as opposed to directly binding to the pre-tRNA binding sites on *At*PRORP1 to inhibit pre-tRNA binding. Anion sites on proteins have also been probed by comparing the dependence of binding affinity on the concentrations of monovalent (M^+) and divalent (M^{2+}) cations for a given anion (deHaseth *et al.* 1977; Barkley *et al.* 1981). For a protein binding to dsDNA in the absence of specific anion binding sites, the theoretical ϕ^{Mg}/φ^{Na} is 0.53, which corresponds to the difference in the cations' occupancy on the phosphodiester bonds in the backbone (deHaseth *et al.* 1977). To test whether anion binding to *At*PRORP1 contributes to the salt dependence of binding affinity, we measured dissociation constants in the presence of varying concentrations of $CaCl_2$ (alone) or Na_2SO_4.

Table 2-2 Effects of anion identity on affinity for pre-tRNA^{Asp}.

Na^+ Salt[a]	K_D (nM)[b]	Lyotropic series
Ac^-	32 ± 7	
SO_4^{2-}	70 ± 10	More kosmotropic
Cl^-	155 ± 20	
Br^-	450 ± 160	
NO_3^-	490 ± 60	
I^-	2900 ± 500	More chaotropic

a: Supplied as the sodium salt at 330 mM Na^+.
b: Value and error reported are from fitting a hyperbola to the results of two independent experiments plotted together.

The pre-tRNA^{Asp} anisotropy in the absence of PRORP and Na^+ displays a hyperbolic dependence on $CaCl_2$ ($K_{D,app}$ = 11 ± 3 mM, Figure 2-3A). This behavior is likely due to stabilization of the tRNA structure upon addition of divalent cations (Leroy *et al.* 1977), as the 5'-fluorescein likely has less rotational freedom when the tRNA structure is compact (Liu *et al.* 2014). The anisotropy of free pre-tRNA increased less than 20% above 50 mM Ca^{2+}, so we varied the $CaCl_2$ concentration from 50 to 125 mM and observed a decrease in *At*PRORP1 affinity (Figure 2-3B). Fitting Equation 4 to the data with Z = 4, we obtained ϕ^{Ca} = 0.51 ± 0.04. This value is in relatively good agreement with ϕ^{Mg} = 0.47 for dsDNA (deHaseth *et al.* 1977). For the PRORP/pre-tRNA complex, the ratio of the slopes for the dependence of log K_D on log concentration (Ca^{2+}/Na^+) is 0.54, similar to the Mg^{2+}/Na^+ ratio of 0.53 for protein/DNA (deHaseth *et al.* 1977). Thus, the salt-dependence of PRORP binding affinity can be explained using only the occupancies of the cations on backbone phosphodiester bonds and excluding anions. The slope of –log(K_D) as a function of log[Na_2SO_4] is smaller than the slope of the NaCl data, resulting in tighter binding at

higher Na^+ concentrations in Na_2SO_4 (Figure 2-3B). However, when –log(K_D) is plotted against ionic strength (–log[I]), the NaCl and Na_2SO_4 data nearly overlay, while the $CaCl_2$ data remains distinct (Figure 2-3C). Thus, the modest differences in affinity observed in NaCl compared to Na_2SO_4 are most likely due mainly to non-specific screening by the ionic atmosphere, to which Na_2SO_4 contributes less due to lower concentration of SO_4^{2+} at a given concentration of Na^+. Given these data, we exclude the term for anion binding sites from our fits. The 1 M NaCl, 1 M Na_2SO_4, and 1 M $CaCl_2$ ionic strength intercepts are within error (Figure 2-3C), confirming that the non-ionic contributions to binding are ion-independent.

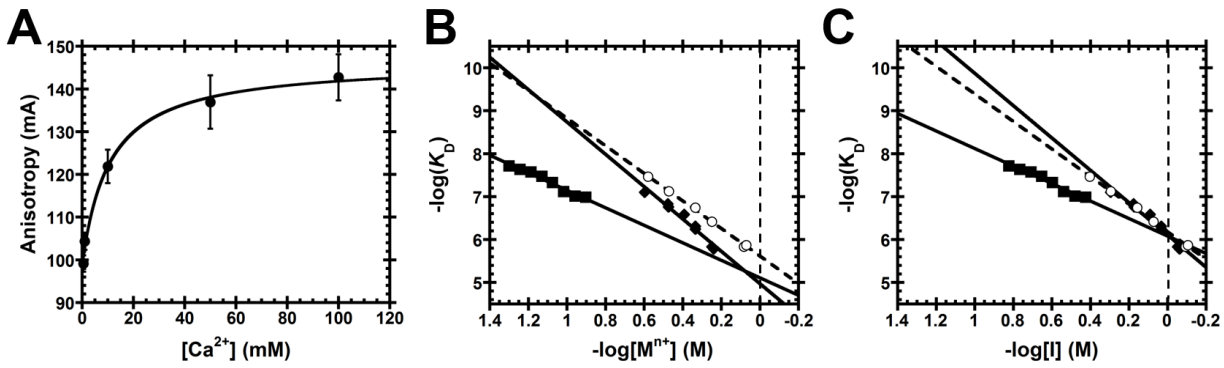


Figure 2-3 Cation (M^{n+}) dependence of dissociation constants for *At*PRORP1 binding to *B. subtilis* fluorescein-labeled pre-$tRNA^{Asp}$.

(**A**) Anisotropy of *Bacillus subtilis* pre-$tRNA^{Asp}$ in the absence of PRORP is dependent on $CaCl_2$ concentration. Data reported as the mean and standard deviation of four independent experiments. A hyperbola (Equation 1, Materials and Methods) was fit to the data ($K_{D,app}$ = 11 ± 3 mM). (**B**) Equation 4 (Materials and Methods) was fit to the data with the dsDNA φ^{Na} = 0.88 or ϕ^{Mg} = 0.47. Data include the dependence of *At*PRORP1 affinity on: NaCl (♦, Z = 4.3 ± 0.3, log(K_0) = 5.0 ± 0.1), Na_2SO_4 (○, Z = 3.6 ± 0.1, log(K_0) = 5.62 ± 0.03), and $CaCl_2$ (■, Z = 4.3 ± 0.3, log(K_0) = 5.1 ± 0.2). The slope of the line ($Z\varphi$ or $Z\phi$) reports on the apparent number of ionic interactions made to substrate phosphodiester bonds, while the intercept [log(K_0)] reports on the non-ionic contributions to affinity. (**C**) Ionic strength (I) dependence of *At*PRORP1 binding to pre-$tRNA^{Asp}$, plotted as the –log(K_D) versus –log[I]. Equation 4 (Materials and Methods) was fit to the data with φ^{Na} = 0.88 or ϕ^{Ca} = 0.47. Data include *At*PRORP1 binding in NaCl (♦; log(K_0) = 6.11 ± 0.04), Na_2SO_4 (○; log(K_0) = 6.18 ± 0.02), and $CaCl_2$ (■; log(K_0) = 6.08 ± 0.09).

***At*PRORP1 makes fewer contacts to substrate leader than the bacterial ribozyme**

Varying the leader length of pre-tRNA substrates beyond 1-nt (Howard *et al.* 2016) or 2-nt (Brillante *et al.* 2016) was previously shown to have little effect on the single-turnover activity and binding affinity with *At*PRORPs. From these data, it is apparent that *At*PRORPs can process a substrate with short 1- and 2-nt leaders, and that *At*PRORP1 discriminates against binding the tRNA product (> 30-fold lower affinity for tRNA than pre-tRNA). In contrast, the *B. subtilis* RNA-based RNase P relies on extensive contacts with the leader and trailer sequences for substrate recognition and displays a significant dependence on leader length beyond 2 nt (Crary *et al.* 1998; Rueda *et al.* 2005).

We determined the Na^+-dependence of affinity for the fluorescein-labeled 1-nt pre-$tRNA^{Asp}$ and $tRNA^{Asp}$ product to evaluate the nature of the *At*PRORP1 interactions with the leader (Table 2-1, Figure 2-4). For the 1-nt substrate, the main effect is a value for K_0 that is increased 2-fold compared to the 5-nt substrate, suggesting a 0.5 kcal/mol reduction in non-ionic interactions with the shorter leader. The value of Z is also reduced modestly, although not enough to indicate the full loss of a phosphodiester contact. In contrast, the Z value for $tRNA^{Asp}$ is reduced to 2.8 ± 0.5, consistent with the loss of one full phosphodiester bond contact. PRORPs making contact to a phosphodiester bond in the 5' leader would be a feature of recognition in common with bacterial RNase P, which contacts the N_{-3}/N_{-2} phosphodiester bond (Hansen *et al.* 2001). The value of K_0 also increases significantly compared with pre-$tRNA^{Asp}$ containing either a 5 nt or 1 nt leader, equivalent to a loss of 2.1 and 1.6 kcal/mol, respectively, indicative of non-ionic interactions with the leader. These data demonstrate that non-ionic interactions with the tRNA body are important

determinants of binding affinity, contributing 4.8 kcal/mol, or ~ 70% of the non-ionic binding energy (Table 2-1).

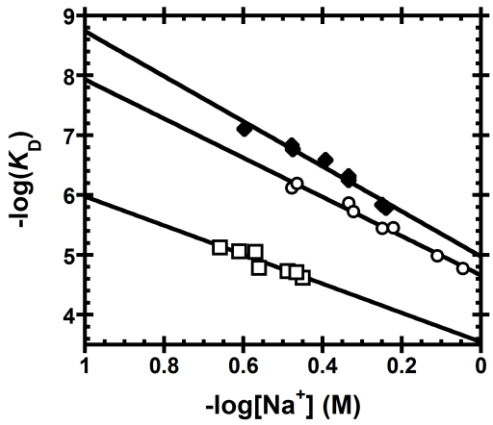


Figure 2-4 Sodium dependence of *At*PRORP1 affinity for substrates with varied leader lengths.

Data include *At*PRORP1 binding to 5' fluorescein-labeled substrates, including: 5-nt pre-tRNA^{Asp} (♦, Z = 4.3 ± 0.3, log(K_0) = 5.0 ± 0.1), 1-nt pre-tRNA^{Asp} (○, Z = 3.7 ± 0.2, log(K_0) = 4.7 ± 0.1), and tRNA^{Asp} (□, Z = 2.8 ± 0.5, log(K_0) = 3.5 ± 0.2). Equation 4 (Materials and Methods) was fit to the data. The slope of the line (Z) reports on the apparent number of ionic interactions made to substrate phosphodiester bonds, while the intercept [log(K_0)] reports on the non-ionic contributions to affinity.

The PRORP PPR domain recognizes tRNAs using non-canonical positions

Previous work demonstrated that mutations of N136T, T180N, and G215N, each at position 6 of an *At*PRORP1 PPR motif, resulted in minor pre-tRNA processing defects (Imai *et al.* 2014). Likewise, T113S and T113N mutations in *At*PRORP2, equivalent to T180 in *At*PRORP1, resulted in little or no processing defects (Brillante *et al.* 2016). To further characterize substrate recognition by the PRORP PPR domain we measured the pre-tRNA binding affinity and salt-dependence for variants of seven residues in *At*PRORP1. The targeted side chains are: (1) highly- or fully-conserved among plant PRORPs, as judged from alignment of PRORPs from 16 species; (2) located on the PPR surface facing the NYN domain; and (3) have the potential to make hydrogen-bonding, ionic, or base-stacking interactions (Figure 2-5). In addition to the residues at

the 1, 3, and 6 positions that have been shown to be involved in base selection in other PPR domains (Barkan *et al.* 2012; Yagi *et al.* 2013), we also targeted residues at position 10, which was not identified in the canonical base-selection motifs (Kobayashi *et al.* 2012). While residues at position 3 in ssRNA binding PPR proteins are typically hydrophobic (i.e. Leu, Phe), in PRORPs the residues at this position are mostly small or hydrophilic. Figure 2-5A shows the position of the residues that we targeted: Y133 (position 3; PPR2), N136 (position 6; PPR2), Y140 (position 10; PPR2), N175 (position 1; PPR3), T180 (position 6; PPR3), R184 (position 10; PPR3), and R212 (position 3; PPR4). We examined the effects of alanine mutations at each position, as well as more conservative mutations including Y133F, T180S, and R184K.

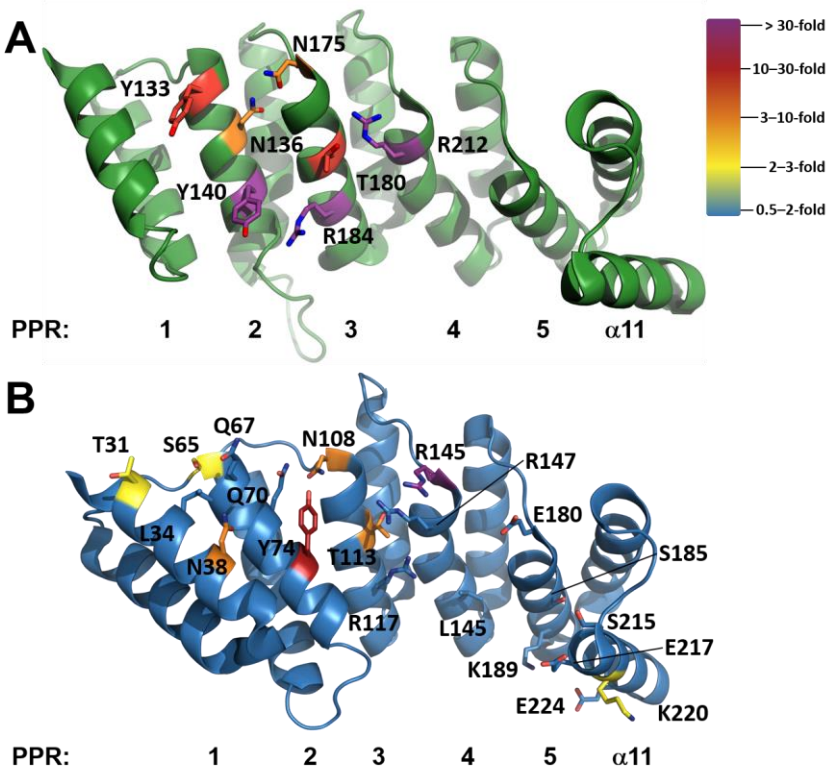


Figure 2-5 Residues selected for mutation in the PPRs of *At*PRORPs, generated in PyMOL

(**A**) *At*PRORP1 (PDB 4g24) PPRs are numbered left (PPR1) to right (helix α11). Residues that were targeted for mutation are numbered. Carbon atoms are color-coded by the largest effect on binding as indicated by the color ramp. (**B**) AtPRORP2 (PDB 5diz) PPRs are numbered left (PPR1) to right (helix α11). Residues that were targeted for mutation are numbered. Carbon atoms are color-coded as in **A**. Alanine mutants were not soluble for Q70, R117, and R147, so no data were collected for these variants. Structure was generated by Schrödinger et al. 2010.

We measured the binding affinities of the *At*PRORP1 mutants for the *B. subtilis* 5-nt pre-$tRNA^{Asp}$ substrate using the FA assay at 330 mM NaCl; these data are summarized in Table 2-3. We observed the largest reductions in binding affinity for the Y140A and R184A variants with decreases of > 190- and 67-fold, respectively. Representative binding data for these variants at 330 mM NaCl can be found in Figures 2-6A and 2-6B. T180 was the only residue in a canonical PPR base-selection position that we tested with a strong effect on binding; T180A reduces the binding affinity by approximately 20-fold compared to WT *At*PRORP1. The other canonical base-selecting residues that we mutated led to modest decreases in binding affinity: N136A had a 6.3-fold effect and N175A had a 9-fold effect. The final residue examined, R212A, eliminated binding as measured with the anisotropy assay (K_D > 30 μM) (Figure 2-7B). Additionally, the enzymatic activity in an STO assay remained lower than the wild-type value even with > 35 μM enzyme and high Mg^{2+} concentrations and several significant miscleavage bands were observed (Figure 2-7A).

We parsed the determinants of substrate binding by *At*PRORP1 in more detail by analyzing the salt dependence of the mutants. In general, the mutations had little effect on the *Z* value for the Na^+-dependence of binding affinity, but they affected the intercept value, K_0 (Table 2-3). These results indicate that the mutated side chains do not form ionic interactions with the phosphodiester backbone of pre-tRNA, rather mediating non-ionic interactions with the substrate. The largest measurable reduction in affinity (> 190-fold) was observed for the Y140A variant, corresponding to a loss of 2.8 kcal/mol of non-ionic binding energy. However, the Y140F mutation only increased K_0 by 6.5-fold, corresponding to a loss of approximately 1 kcal/mol in non-ionic interactions compared to WT *At*PRORP1. These results are consistent with PRORP interacting with pre-tRNA with both the tyrosine hydroxyl and the phenyl ring (Guckian *et al.* 2000). For the R184A mutant,

the 67-fold reduced affinity corresponds to a reduction of 2.5 kcal/mol compared to WT, while the R184K mutation increased the K_0 by 10-fold, corresponding to a loss of 1.4 kcal/mol of binding energy compared to WT.

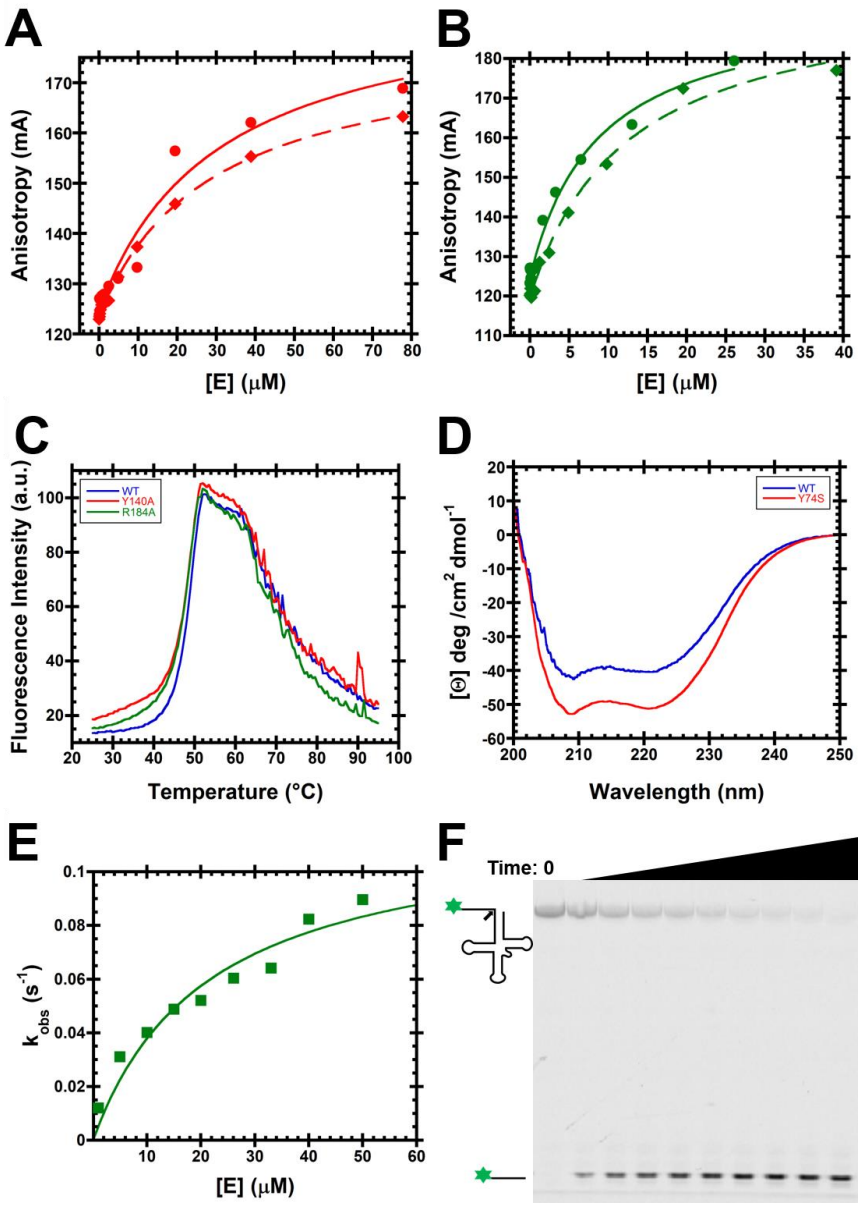
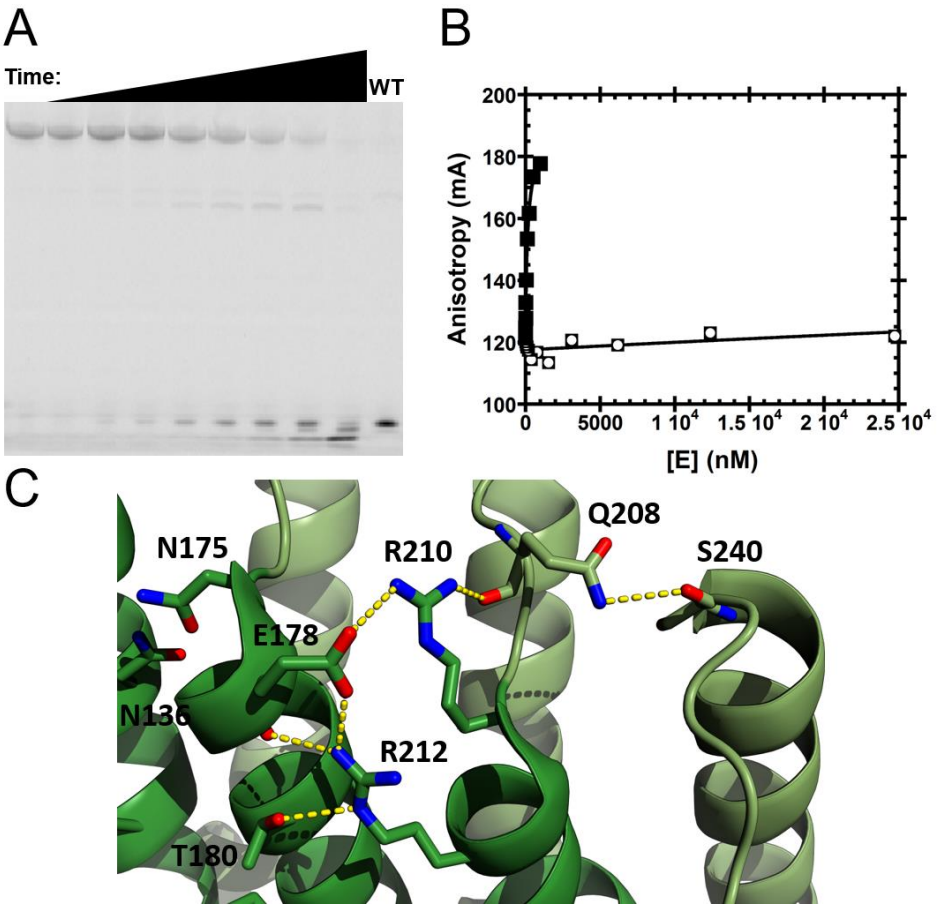


Figure 2-6 Stability of Y140A and R184A AtPRORP1 or Y74S AtPRORP2.

(**A**) Binding curves for Y140A AtPRORP1 collected at 330 mM NaCl, as described in the Materials and Methods. (**B**) Binding curves for R184A AtPRORP1 collected at 330 mM NaCl, as

described in the Materials and Methods. (**C**) Thermofluor melting curves for WT, Y140A, and R184A *At*PRORP1, collected and analyzed as described in the Materials and Methods. The T_m values are 48.3 ± 0.1 (WT), 47.7 ± 0.2 (Y140A), and 47.4 ± 0.2 (R184A). (**D**) Circular dichroism



data for WT and Y74S *At*PRORP2, collected and analyzed as described in the materials and methods. (**E**) Single turnover kinetic analysis of R184A *At*PRORP1 with B. subtilis pre-$tRNA^{Asp}$, collected at 25°C in 30 mM MOPS, pH 7.8, 150 mM NaCl, 20 mM $MgCl_2$, and 1 mM TCEP. Fitting a hyperbola to the data, we obtain k_{max}^{E} = 0.12 ± 0.01 s^{-1} and $K_{1/2}^{E}$ = 21 ± 7 μM. The WT enzyme value under these conditions is ~ 0.1 s^{-1} (Howard et al. 2015). (**F**) A representative gel from the experiments described in panel E, at 40 μM enzyme. Time points (left to right) are 0, 6 sec, 16 sec, 26 sec, 40 sec, 1 min, 2.5 min, 5 min, 10 min, and 30 min.

Figure 2-7 Defects in R212A *At*PRORP1 catalysis and substrate binding with *B. subtilis* pre-$tRNA^{Asp}$.

(**A**) Single turnover assay with 37 μM R212A and 30 nM substrate. Time points (left to right) are 0, 1 min, 5 min, 10 min, 30 min, 1 hr, 2 hr, 4 hr 20 min, and 20 hr 40 min. 5 μM WT enzyme at 30 minutes (at right) was included for comparison. (**B**) Anisotropy pre-$tRNA^{Asp}$ binding to R212A (◘) or WT (■) *At*PRORP1. A hyperbola (Equation 1, Materials and Methods) was fit to the data ($K_{D,app}$ > 300000 nM; lower limit estimated assuming the change in anisotropy for fully-bound substrate is the same as WT). (**C**) Arg 212 in *At*PRORP1 takes part in an extended series of interactions involving both main chain and side chain atoms, beginning with Thr 180 (PPR 3 helix

A) and terminating with Ser 240 (PPR 4 helix B). Generated in PyMOL (PDB 4g24) (Schrödinger 2010).

Table 2-3 Na^+-dependence of *At*PRORP1 variants affinity for *B. subtilis* pre-$tRNA^{Asp}$.

***At*PRORP2 residue**[a]	***At*PRORP1 Variant**	**K_D (nM)**[b]	**Fold-WT**	**Z**[c]	**–log(K_0)**[c]	**ΔG_0 (kcal/mol)**[d]
–	WT	155 ± 20	1.0	4.3 ± 0.3	5.0 ± 0.1	–6.9 ± 0.1
Q67	Y133A	2600 ± 200	17	4.4 ± 0.3	3.8 ± 0.2	–5.2 ± 0.3
	Y133F	4600 ± 300	30	4.2 ± 0.2	3.7 ± 0.1	–5.1 ± 0.1
Q70	N136A	980 ± 180	6.3	3.9 ± 0.2	4.3 ± 0.1	–6.0 ± 0.1
Y74	Y140A	29700 ± 7000	192	4.2 ± 0.4	3.0 ± 0.3	–4.1 ± 0.4
	Y140F	1000 ± 200	6.5	4.1 ± 0.6	4.3 ± 0.3	–5.9 ± 0.4
N108	N175A	1400 ± 400	9.0	3.8 ± 0.2	4.4 ± 0.2	–5.9 ± 0.3
T113	T180A	3300 ± 1500	22	4.0 ± 0.2	3.8 ± 0.1	–5.2 ± 0.1
	T180S	1700 ± 300	11	4.6 ± 0.4	3.9 ± 0.2	–5.4 ± 0.3
R117	R184A	10400 ± 2400	67	4.5 ± 0.5	3.2 ± 0.3	–4.4 ± 0.3
	R184K	1900 ± 300	12	4.6 ± 0.5	4.0 ± 0.3	–5.5 ± 0.4
R147	R212A[e]	> 30000	> 194	ND[f]	ND	ND

a: Amino acid in *At*PRORP2 that is in the homologous position to the side chain in *At*PRORP1.
b: Value and error reported are from fitting a hyperbola to the results of two independent experiments in 330 mM NaCl plotted together.
c: Value and error from fitting Equation 4 to the log-log plot of the Na^+-dependence data using φ^{Na} = 0.88 as described in the Materials and Methods.
d: Calculated using $\Delta G_0 = -RT \times \ln K_0$.
e: Affinity for the R212A mutant was not measurable; little change in anisotropy was observed at 25 μM *At*PRORP1 (Figure 2-8B).
f: Not determined (ND).

The Y133 variants revealed a relationship different from that of the Y140 variants. The Y133F and Y133A variants reduced pre-tRNA affinity by comparable values, 1.8 kcal/mol and 1.6 kcal/mol, respectively. These data suggest that the hydroxyl group, but not the phenyl ring of Y133, contributes to substrate affinity. The N136A and N175A mutations reduce the value of the intercept corresponding to approximately the loss of 1 kcal/mol in non-ionic binding energy apiece. We estimated comparable free energy losses for T180A and T180S variants, 1.7 and 1.5 kcal/mol, respectively, compared to that of WT.

The *At*PRORP2 PPR domain recognizes tRNAs using a similar binding surface

To determine whether this binding surface is shared among *Arabidopsis* PRORPs and whether additional PRORP PPR side chains are important for substrate binding, we screened the pre-tRNA binding affinity of 22 alanine variants in *At*PRORP2 (Figure 2-5B). Previously, we and others proposed that the 1st, 3rd, 6th and 10th residues (numbered as in Barkan, *et al*. 2012) in each PPR motif could periodically contribute to substrate binding in *At*PRORPs (Figure 2-8) (Barkan *et al.* 2012; Kobayashi *et al.* 2012; Yin *et al.* 2013; Karasik *et al.* 2016). Therefore, we systematically targeted residues in these positions for all five PPR motifs and the PPR C-terminal helix (α11) for alanine mutagenesis. This analysis necessarily excludes three alanine residues (A110, A150, and A182).

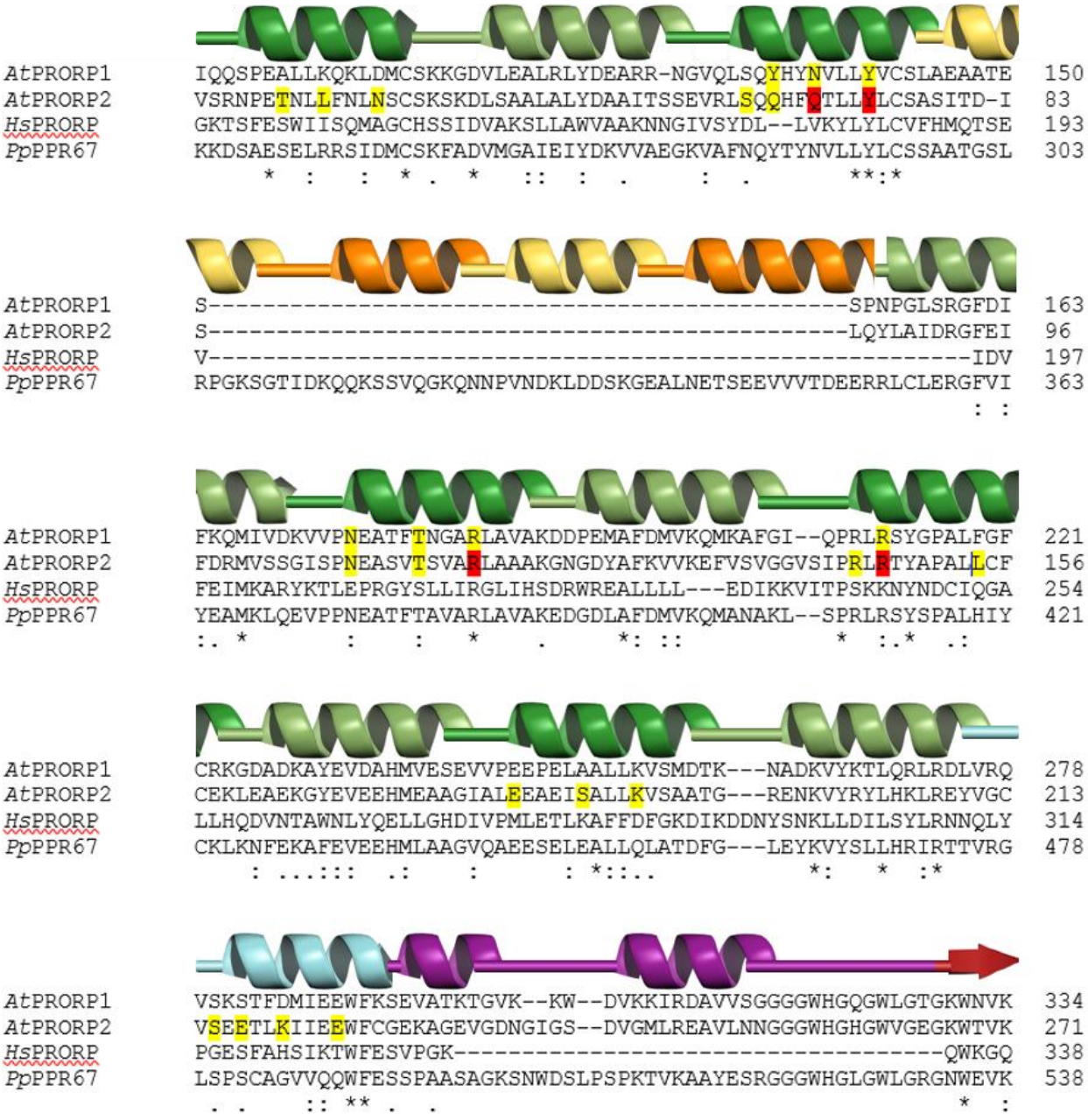


Figure 2-8 Alignment of PRORP PPR domains (Clustal Omega) (Goujon et al. 2010; Sievers et al. 2011).

The three Arabidopsis thaliana PRORPs (*At*PRORP1-3) were aligned to the human PRORP (HsPRORP) and PRORPs from the moss Physcomitrella patens (PpPPR63, PpPPR67, and PpPPR104). For clarity, *At*PRORP3, PpPPR67, and PpPPR104 are excluded from the figure. The boundaries of additional PPR helices in P. patens PRORPs were assigned using models generated with *At*PRORP1 (PDB 4g23) and chloroplast PPR protein 10 (PDB 4m59) as templates in the SWISS-MODEL ExPASy server (Arnold et al. 2006; Guex et al. 2009; Kiefer et al. 2009; Biasini et al. 2014). PPR A helices, which were targeted for mutation, are colored dark green, while PPR B helices are colored pale green. Additional putative PPR A helices in PpPPR67 are colored orange, while putative PPR B helices are colored beige. PPR helix 11 (numbered as in *At*PRORPs)

is colored cyan, while a plant-specific helical insert is purple and the first β-strand of the central domain is red. *At*PRORP1 and *At*PRORP2 residues mutated in this manuscript are highlighted in yellow. *At*PRORP2 residues for which alanine mutants were insoluble are highlighted in red.

We examined the binding affinity of the alanine mutants with *A. thaliana* nuclear 6-nt pre-tRNA^{Gly} and *B. subtilis* 5-nt pre-tRNA^{Asp} substrates. The residues that we identified with alanine scanning mutagenesis of *At*PRORP2 are consistent with the binding surface identified in *At*PRORP1. The Q67A (position 3; PPR2), N108A (position 1; PPR3), T113A (position 6; PPR3), and R145A (position 1; PPR4) mutants increased the K_D values for pre-tRNA by at least 1.5-fold in the highest fold-increase that we observed compared to wild type *At*PRORP2 (Table 2-4). Importantly, these residues (excluding R145) correspond to three of the seven positions we identified in the *At*PRORP1 PPR domain. This analysis also identified additional residues that alter pre-tRNA binding affinity beyond those evaluated in *At*PRORP1, all of which fall primarily within the nearby PPR surface. The K_D values for pre-tRNA of the N38A (position 10; PPR1), S65A (position 1; PPR2), T31A (position 3; PPR1), and K220A (position 6; α11) mutants increased by at least 2-fold in the highest fold-increase that we observed compared to wild type *At*PRORP2. Four *At*PRORP2 alanine mutants – Q70A (position 6; PPR2), Y74A (position 10; PPR2), R117A (position 10; PPR3), R147A (position 3; PPR4) – did not express as soluble proteins, suggesting that mutation of these residues may affect the stability of *At*PRORP2. These residues are all located on the proposed substrate-binding surface, and include four of the seven residues that alter substrate affinity in *At*PRORP1.

Given the importance of Y140 in *At*PRORP1 for pre-tRNA affinity, we further investigated the type of interaction between this amino acid and pre-tRNA by analyzing the Y74S and Y74F mutants in *At*PRORP2. We found that these mutations significantly decreased substrate affinity (Table 2-4), as observed for *At*PRORP1. These results further demonstrate that Y140 interacts with

substrate using both the phenyl ring and the hydroxyl group. Taken together, the above results indicate that surface we identified in *At*PRORP1, primarily in PPR2 and PPR3, generalizes as the major surface involved in PRORP substrate binding.

Table 2-4 Effects of mutation to *At*PRORP2 PPRs on affinity for pre-tRNA.

***At*PRORP1 Residue**	***At*PRORP2 Residue**	**Position (PPR/Residue)**	**pre-tRNA^{Gly}**		**pre-tRNA^{Asp}**	**Fold-WT**
			K_D **(nM)**[a,b]	K_D **(nM)**[a,c]	K_D **(nM)**[a,b]	
–	WT	–	58 ± 8	260 ± 30	24 ± 5	1.0
A98	T31A	1 / 3	60 ± 8	530 ± 50		2.0
K101	L34A	1 / 6	28 ± 6	170 ± 10		0.7
D105	N38A	1 / 10	160 ± 40	1410 ± 150	70 ± 20	5.4
S131	S65A	2 / 1	100 ± 20	590 ± 80	35 ± 8	2.3
Y133	Q67A	2 / 3	70 ± 10	390 ± 70	30 ± 5	1.5
Y140	Y74S	2 / 10	1340 ± 170		110 ± 30	23
	Y74F		170 ± 20		45 ± 9	2.9
N175	N108A	3 / 1	190 ± 30	720 ± 50	70 ± 20	3.3
T180	T113A	3 / 6	70 ± 20	900 ± 130	30 ± 3	3.5
R210	R145A[d]	4 / 1	> 2000			> 34
F119	L154A	4 / 10	60 ± 10	230 ± 20		1.0
E245	E180A	5 / 1	35 ± 4	150 ± 20		0.6
A250	S185A	5 / 6	40 ± 6	350 ± 40		1.3
K254	K189A	5 / 10	52 ± 9	350 ± 30		1.3
S280	S215A	α11 / 1	66 ± 21	360 ± 60		1.4
S282	E217A	α11 / 3	56 ± 14	260 ± 20		1.0
D285	K220A	α11 / 6	129 ± 64	250 ± 40		2.2
E289	E224A	α11 / 10	55 ± 7	310 ± 35		1.2

a: Value and error reported are from fitting a hyperbola to the results of three independent experiments plotted together.

b: Measured in 30 mM MOPS pH 7.8, 1 mM DTT, 150 mM NaCl, and 6 mM $CaCl_2$.
c: Measured in 30 mM MOPS pH 7.8, 1 mM DTT, 330 mM NaCl, and 6 mM $CaCl_2$.
d: Low protein yield for R145A limited us to two binding assays, neither of which reached saturation up to 7500 nM. A lower limit for the K_D was estimated from the available data.

PRORP PPR domain appears not to recognize tRNAs with base-specificity

To investigate whether *At*PRORPs recognize pre-tRNA in a base-specific manner, we mutated residues that the previously established PPR recognition codes suggest should recognize pyrimidines (Barkan *et al.* 2012; Yagi *et al.* 2013). *At*PRORP1 contains amino acids located between PPR motifs 2 and 3 (Y133/N136/N175) that should recognize pyrimidines (Barkan *et al.* 2012; Yagi *et al.* 2013). For this analysis, we examined pyrimidines in pre-tRNA that would likely interact with PRORPs. The D- and TψC-loops (tRNA elbow) in pre-tRNA have been proposed to interact with the PPR domain (Gobert *et al.* 2013). We assumed that the pyrimidines should not be in secondary/tertiary contacts or otherwise buried and inaccessible in the unbound tRNA structure, limiting the proposed interaction to uridines at positions 16, 17, 20, and 21 in pre-$tRNA^{Asp}$ (Figure 2-2A). Mutation to adenosine at each of these positions altered the affinity by at most 2-fold (Table 2-5), demonstrating a lack of sequence-specific interaction at these sites. In combination with the previous data regarding the effects of tRNA mutations on PRORP binding/catalysis (Imai *et al.* 2014; Brillante *et al.* 2016), these data reinforce the hypothesis that PRORPs utilize a mode of substrate recognition different from the previously described PPR base-selection determined in ssRNA binding proteins. However, the possibility remains that there are other sites in pre-tRNA that interact with PRORP PPRs in a sequence-specific manner.

Table 2-5 Effects of *B. subtilis* pre-tRNA^{Asp} mutants on affinity for *At*PRORP1.

Variant	K_D (nM)[a]	Fold-WT
WT	155 ± 20	1
U16A	140 ± 20	0.90
U17A	100 ± 10	0.65
U20A	200 ± 10	1.29
U21A	300 ± 30	1.94

a: Value and error reported are from fitting a hyperbola to the results of two independent experiments in 330 mM NaCl plotted together.

Na^+ screening inhibits *At*PRORP1 single-turnover activity

While it is possible that the ionic strength affects only the binding affinity, it might also affect other aspects of PRORP catalysis. To determine whether the NaCl concentration affects cleavage catalyzed by PRORP, we performed single-turnover (STO) activity assays with 5 μM *At*PRORP1, which is saturating under low NaCl conditions, and limiting (30 nM) *B. subtilis* pre-tRNA^{Asp}. We used concentrations of $MgCl_2$ that we previously determined to be either saturating (20 mM) or sub-saturating (1.25 mM) for catalysis (Howard *et al.* 2015). The observed rate constant (k_{obs}) is independent of the NaCl concentration between 25 mM and 200 mM, but is reduced at higher NaCl concentrations (Figure 2-9). The concentration dependence of NaCl inhibition above 400 mM is similar for saturating and sub-saturating $MgCl_2$. Fitting a general inhibition model (Equation 3, weighted fit, Materials and Methods) with a variable Hill Coefficient (n^{Na}) to the data yields similar IC_{50} (310 ± 70 mM for 20 mM $MgCl_2$ and 360 ± 70 mM for 1.25 mM $MgCl_2$) and n^{Na} values (4.5 ± 1.3 for 20 mM $MgCl_2$ and 5.0 ± 1.3 for 1.25 mM $MgCl_2$) for

both $MgCl_2$ conditions (Figure 2-9). The Hill coefficients for STO inhibition (4.5–5) are in good agreement with the cooperativity we observe for inhibition of pre-tRNA binding.

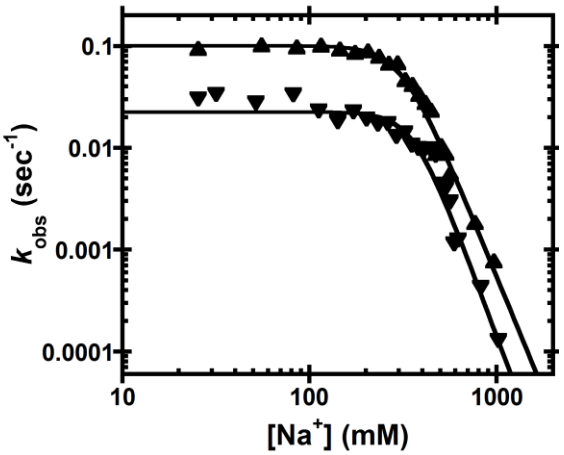


Figure 2-9 Na^+-dependence of *At*PRORP1 cleavage activity.

The dependence of the *At*PRORP1-catalyzed cleavage of fluorescein-labeled pre-$tRNA^{Asp}$ on the [NaCl] was measured under single turnover conditions (k_{obs}). Data include *At*PRORP1-catalyzed cleavage in 20 mM (▲) or 1.25 mM (▼) $MgCl_2$. Equation 3 (Materials and Methods, weighted fit) was fit to the data. For 20 mM $MgCl_2$, IC_{50} = 310 ± 70, n^{Na} = 4.5 ± 1.3. For 1.25 mM $MgCl_2$, IC_{50} = 360 ± 70, n^{Na} = 5.0 ± 1.3.

If the observed inhibition is due to decreased affinity of *At*PRORP1, the activity should begin to decrease at NaCl ~ 530 mM (the point at which [E] becomes < 5 times the K_D, as measured in Ca^{2+}), assuming that the binding affinity is equivalent in Ca^{2+} and Mg^{2+}. However, the activity is inhibited at lower NaCl concentrations (~ 300 mM). To test whether this observation is only due to an effect of NaCl on binding affinity, as opposed to NaCl affecting other steps in catalysis, we measured the STO k_{obs} at 20 mM $MgCl_2$ and 350 mM NaCl with varying the enzyme concentration (data not shown). The STO k_{obs} is strongly dependent on enzyme concentration under these conditions, indicating that 5 μM PRORP1 is sub-saturating, thus requiring that the STO $K_{1/2}$ in $MgCl_2$ is > 30-fold higher than the thermodynamic K_D (155 ± 20 nM) in $CaCl_2$. These data also indicate that PRORP-substrate affinity is reduced in $MgCl_2$ as compared to $CaCl_2$. This cation-

dependent decrease in affinity is consistent with an EcoRV catalytic mutant that binds its cognate DNA ≈ 40-fold weaker in Mg^{2+} than in Ca^{2+} (Martin *et al.* 1999). In summary, the STO data are consistent with NaCl disrupting PRORP-substrate binding, but not other kinetic steps.

2.5 Discussion

The goal of this work was to characterize the molecular interactions between *At*PRORPs and pre-tRNA using the ion-dependence of binding and catalysis and mutations in the PPR domain. The NaCl-dependence of pre-tRNA affinity of *At*PRORPs revealed 4–5 interactions with phosphodiester bonds in pre-tRNA and approximately 6 to 7 kcal/mol of non-ionic binding energy. These narrow ranges support previous observations that pre-tRNA recognition by *At*PRORPs is similar across paralogs and largely independent of sequence. Depending on the context, biological hydrogen bonds can supply 0.5–3 kcal/mol of free energy (Fersht *et al.* 1985), while van der Waals interactions such as base stacking can supply 0.5–2 kcal/mol (Guckian *et al.* 2000). *At*PRORPs likely form a combination of hydrogen bonds and van der Waals interactions with pre-tRNA.

Comparison of the Na-dependence and mutation of residues in the substrate binding domain in both *At*PRORP1 and 2 indicate a similar binding surface with modest differences in substrate recognition, as previously evaluated (Karasik *et al.* 2016). In general, alanine mutations affected the substrate affinity of *At*PRORP1 to a greater extent than *At*PRORP2. This might be due to the lower $CaCl_2$ conditions used for affinity measurements with *At*PRORP2, which also reduced the WT *At*PRORP1 binding affinity (data not shown). The variety of ion-dependent effects on PRORP binding affinity that we observe can be synthesized as follows.

Cations such as Na^+ and Ca^{2+} inhibit PRORP/pre-tRNA binding by interacting with and competing for phosphodiester bonds on pre-tRNA. We observe a log-linear decrease in PRORP/pre-tRNA binding affinity with respect to [Na^+] above 180 mM NaCl (Figures 2-2C, 2-

3B, 2-3C, and 2-4). The PPR mutations we generated (described in further detail below), do not affect the slope of the Na^+-dependence of binding at high [Na^+] (Table 2-3). However, we observe lesser effects from these mutations on binding affinity below 180 mM NaCl (Tables S1 and S3). While Ca^{2+} alone also inhibits PRORP/pre-tRNA binding (Figures 2-3B and 2-3C), we observe a positive effect on *At*PRORP1/pre-tRNA binding with high [Na^+]. Overall, these data are most informative with respect to the differences we observe between WT and variant PRORPs and pre-tRNAs, which allow us to parse the interactions between various aspects of each.

PRORP PPR domain

Our extensive mutagenesis data in the PPR domain allow us to contrast PRORP-RNA binding with previously described PPR proteins. PPR-containing proteins are a large family with the structurally conserved ≈ 35 residue helix-turn-helix motif found in tandem repeats that have been implicated in RNA metabolism (Small and Peeterns 2000; Schmitz-Linneweber and Small 2008). PPR proteins are found broadly in eukaryotes, with land plants having the largest set of PPR proteins. For example, there are over 400 members in *Arabidopsis* and *Oryza* (O'Toole *et al.* 2008). Some PPR proteins bind target RNAs in a sequence-specific manner, with recognition of a nucleobase achieved primarily by residues 6 and 1' (also numbered 4 and ii by Yagi, *et al*., or 5 and 35 by Yin, *et al*.) on the A helices of two tandem PPR motifs (Figure 1B, Figure S1; A helices colored dark green) (Barkan *et al.* 2012; Yagi *et al.* 2013; Yin *et al.* 2013). In some cases, the binding sites have been identified in the UTRs of mRNAs, where the PPR proteins are proposed to regulate splicing, translation and/or stability of the mature transcript (Schmitz-Linneweber and Small 2008; Chen *et al.* 2016b). One goal of this work was to determine whether PRORP PPR domains use the same or similar mechanism to recognize pre-tRNAs. The results of our mutagenesis experiments indicate that PRORP PPRs use a mode of RNA recognition that is

different from ssRNA-binding proteins. The PPR domains of *At*PRORP1 and *At*PRORP2 do not rely on a limited number of residues from every repeat, but rather a more extensive surface primarily located in PPR2 and PPR3, which is consistent with previous data indicating impaired activity for ΔPPR2 and ΔPPR3 *At*PRORP1 variants (Imai *et al.* 2014).

We initially hypothesized that PRORP PPRs would not recognize RNA using base selection as observed with ssRNA-binding PPR proteins because the base-specifying residues in PRORPs are frequently noncanonical (e.g. 6 and 1' are not often asparagine, aspartate, or threonine, Figure 2-8). Furthermore, conservation to other PPR domains yielded predictions of only two or three PPR motifs in PRORPs (Holzmann *et al.* 2008; Gobert *et al.* 2010), yet crystal structures revealed five tandem repeats in *At*PRORP1 (Figure 2-1A) (Howard *et al.* 2012). By homology, the metazoan PRORP also contains at least five repeats, the last three of which have been visualized in a crystal structure of human PRORP with an N-terminal deletion (Reinhard *et al.* 2015). Several substrate binding residues that we identified, such as Y140 and R184, are conserved in metazoan PRORPs (Y183 and R218 in humans). Thus, even though the metazoan PRORPs require additional subunits for catalysis, pre-tRNA recognition by the human PRORP PPR domain could use the same interaction surface as plant PRORP PPRs.

The alanine mutations most detrimental to *B. subtilis* pre-tRNA^{Asp} affinity were Y140A in *At*PRORP1 PPR2 (> 190-fold) and R184A in *At*PRORP1 PPR3 (67-fold); the equivalent mutations in *At*PRORP2 (Y74A and R117A) rendered the protein insoluble. Despite these insoluble variants, we find no evidence that the mutants we have analyzed are less stable than the WT enzyme. The melting temperatures (T_m) for Y140A and R184A *At*PRORP1 are not significantly different from WT (Figure 2-6C), while the CD spectra for Y74S *At*PRORP2 does not reveal a significantly different secondary structure from WT (Figure 2-6D). Furthermore, we were able to recover WT-

like STO activity at high concentrations of R184A (Figure 2-6E), with no apparent miscleavage (Figure 2-6F). These data support our conclusions that the PPR mutations primarily reduces the affinity for pre-tRNA, without affecting substrate positioning or protein stability.

Y140A and R184A *At*PRORP1 mutations similarly bind the *A. thaliana* pre-tRNA^{Cys} substrate weaker than WT in 20 mM $CaCl_2$ and 330 mM NaCl conditions (> 93-fold and > 3.8-fold, respectively). Unexpectedly, the effects of the *At*PRORP1 mutations are greater than the 34-fold decrease in binding affinity reported for a Δ245 *At*PRORP1, which fully lacks the first 4 PPR motifs (Howard *et al.* 2012). However, this measurement was carried out at 1 mM $CaCl_2$ and 100 mM NaCl with the *A. thaliana* mitochondrial pre-tRNA^{Cys} substrate. Under these conditions, the individual mutations have little effect on binding affinity (Table 2-6). This effect is likely explained by the increased dependence of the PRORP-substrate affinity on ionic interactions under lower ionic strength, such that individual non-ionic interactions contribute less significantly to the overall affinity.

Table 2-6 Effects of mutations on AtPRORP1 affinity for pre-tRNA

Variant	**pre-tRNA^{Cys}**				**pre-tRNA^{Asp}**	
	K_D (nM)[a]	**Fold-WT**	**K_D (nM)[b]**	**Fold-WT**	**K_D (nM)[a]**	**Fold-WT**
WT	200 ± 40	1	1720 ± 160	1	155 ± 20	1
Y140A[c]	150 ± 20	0.75	>160000	>93	29700 ± 7000	192
R184A	170 ± 40	0.85	>6500	>3.8	10400 ± 2400	67

a: Value and error reported are from fitting a hyperbola to the results of two independent experiments plotted together in 30 mM MOPS pH 7.8, 1 mM TCEP, 100 mM NaCl, and 1 mM $CaCl_2$.

b: Value and error reported are from fitting a hyperbola to the results of two independent experiments plotted together in 30 mM MOPS pH 7.8, 1 mM TCEP, 330 mM NaCl, 20 mM $CaCl_2$.

c: We could not reach saturating enzyme conditions for Y140A binding to pre-tRNA^{Cys} in 330 mM NaCl and 20 mM $CaCl_2$. A lower limit for K_D was estimated from the available data, assuming the saturating anisotropy was the same as for WT.

The loss of 1 kcal/mol that we observe for the Y140F is consistent with Y140 interacting through a hydrogen bond to substrate with the tyrosine hydroxyl. The additional 1.8 kcal/mol loss observed with the alanine substitution is consistent with the energy supplied by stacking a phenyl ring with a nucleic acid base (Guckian *et al.* 2000). For the R184A mutant, the 67-fold reduced affinity corresponds to a loss of 2.5 kcal/mol in non-ionic binding energy compared to WT, which could indicate loss of 1–3 hydrogen bonds with the guanidinium group and/or hydrophobic interactions with the arginine methylene groups. By comparison, the R184K mutation increased the K_0 by 10-fold, corresponding to a loss of 1.4 kcal/mol compared to WT, consistent with the loss of a hydrogen bond. In *At*PRORP1, Y140 and R184 are located at position 10 of neighboring helices forming a non-sequential, structural pair. This YR pair is widely conserved in PRORP PPRs, including metazoan PRORPs (Figure 2-8).

The C-terminal La domain of the telomerase protein p65 also contains a conserved YR structural pair (Y407 and R465), situated on neighboring β-strands, that are important for recognition of the conserved GA bulge in stem IV of the telomerase RNA (Singh *et al.* 2012). Given the significant contribution of the phenyl ring and the guanidinium group revealed by the Y140A/F and R184A/K mutations in *At*PRORP1 and Y74S/F in *At*PRORP2, we propose that these residues make similar interactions with the tRNA elbow, the conserved structural feature that results from the interaction of the D- and TψC-loops, which were previously proposed to interact with PRORPs (Gobert *et al.* 2013). These interactions do not have to be sequence specific, but like the p65 YR pair (Singh *et al.* 2012), they could favor purines such as the conserved $G_{18}G_{19}$ in the D-loop. In fact, when these residues in a canonical pre-tRNA^{Gly} are individually mutated to

adenine, there is a ≈ 4.5-fold (G_{18}) or ≈ 1.5-fold (G_{19}) increase in the STO $K_{1/2}$ (Brillante *et al.* 2016).

In contrast to our Y140 data, the data for Y133 indicate that the hydroxyl group, but not the phenyl ring, contributes to substrate affinity. Mutation of N136 and N175 results in the loss of approximately 1 kcal/mol apiece, consistent with a hydrogen bond from the amide side chain of each. The T180A mutation leads to a 1.7 kcal/mol loss, consistent with up to one or more hydrogen bonds and/or hydrophobic interactions. Our data show that T113A in PRORP2 had at most a 3.5-fold effect on binding affinity. The T180S variant reduces the affinity to a similar extent, indicating a loss of 1.5 kcal/mol, but substitutions of T113S and T113N were not sufficient for significant processing defects in previous studies (Brillante *et al.* 2016). Nonetheless, our data may either indicate that the threonine methyl group is required to sterically position the hydroxyl for substrate interaction or that it makes a hydrophobic contact to substrate.

While *At*PRORP1 T180 (T113 in *At*PRORP2) is at a base-selecting position 6, its corresponding 1' partner in the putative base-selection position would be R210 (R145 in *At*PRORP2). Arginine residues have not previously been identified as base-selecting residues for PPRs (Barkan *et al.* 2012; Yagi *et al.* 2013). Furthermore, the potential nucleobase binding cleft is occluded by an interaction between T180 (T113) and R212 (R147) in the *At*PRORP1 and *At*PRORP2 crystal structures (PDB 4g24 and 5diz) (Howard *et al.* 2012; Karasik *et al.* 2016). The R212A mutant showed severe defects in substrate binding and catalysis. R212 is involved in multiple interactions within an extended series of salt-bridge and hydrogen bonding interactions that begins at T180 in helix A of PPR3 and terminates in helix B of PPR 4 at S240 (Figure 2-7C) (Howard *et al.* 2012), suggesting that the mutation may disrupt pre-tRNA binding by altering the structure of the PPRs. To examine this, we tested the *At*PRORP1 R212A stability by the

ThermoFluor assay (Pantoliano *et al.* 2001), which demonstrated that the observed melting point was unchanged from WT (data not shown). Taken together, our mutagenesis data for *At*PRORP1 and *At*PRORP2 indicate that the PRORP PPR domain does not interact with pre-tRNA in the same manner as ssRNA-binding PPR proteins.

An alternative recognition mode could include structural recognition, like the tRNA elbow recognition employed by the specificity (S) domain of the P RNA subunit in the bacterial ribozyme. The S-domain makes stacking interactions with the conserved G_{19}–C_{56} tertiary interaction between the D- and TψC-loops and sugar face interactions with the conserved G_{53}–C_{61} pair at the end of the TψC arm (Reiter *et al.* 2010). These interactions recognize conserved pre-tRNA structural elements and allow P RNA to recognize the entire set of pre-tRNA transcripts without specificity for the tRNA body sequence. Likewise, we propose that the PPRs of PRORPs recognize tRNA structure using a similar strategy through structure-specific interactions. A similar mode of recognition has been proposed previously (Gobert *et al.* 2013). While our data support a model in which PRORP PPRs recognize the tRNA structure rather than conserved sequences, we cannot yet rule out at least the partial involvement of base-recognition strategies such as those employed by canonical PPR proteins.

PRORP-substrate recognition model

Using our data, as well as previously published data (Gobert *et al.* 2013; Imai *et al.* 2014), we propose a new model to describe how *At*PRORP1 and *At*PRORP2 bind their substrates. This model is predictive and yields specific hypotheses about the PRORP-pre-tRNA complex. A previous attempt to model the complex utilized *At*PRORP1 bound to tRNA with base-specifying interactions between a TψC-loop cytosine and PPR motifs 3 and 4 (Imai *et al.* 2014). To accomplish this, the tRNA coordinates were modified using a tRNA bound to a pseudouridine

synthetase (Hoang and Ferre-D'Amare 2002), which has significant alterations to the TψC-loop bases and backbone. The remaining tRNA coordinates were derived from the bacterial RNase P holoenzyme in complex with a tRNA product (Reiter *et al.* 2010).

For constructing our model, we employed the crystal structures of *At*PRORP1 (PDB 4g24) and *At*PRORP2 (PDB 5diz) (Howard *et al.* 2012; Karasik *et al.* 2016). The two structures are highly conserved, yet there is a significant difference in the angle between specific PPR motifs and between the central domain and the metallonuclease domain; as a consequence, *At*PRORP2 is in a more "open" conformation. We propose that the two distinct structural snapshots may represent two different conformations that potentially play a role in substrate binding (Karasik *et al.* 2016). The more "open" *At*PRORP2 conformation can more readily accommodate tRNA, since the distance between the active site and the proposed substrate binding region in the PPR domain is ~ 50 Å. Therefore, we used the *At*PRORP2 structure to generate a tRNA interaction model and used this model as a template to generate a model for *At*PRORP1-tRNA recognition. There are no crystal structures available for any of the tRNAs used in this study to use for our PRORP-tRNA model. However, since the 3D structure of tRNAs is highly conserved, we chose a tRNA that resembles a canonical eukaryotic tRNA with an available crystal structure (PDB 1ehz) (Shi and Moore 2000) and used this structure with only slight modifications to bases predicted to interact with the protein.

The tRNA substrate can be accommodated well between the metallonuclease and PPR domains in our model (Figure 2-10), with one exception. There is a short PRORP helix (α21) and part of the loop that precedes it near the active site that sterically clashes with the 3' strand of the tRNA acceptor stem. We posit that this NYN helix will adopt a different conformation upon pre-tRNA binding and may be directly involved in recognition. This region could serve as a hinge that

allows or blocks substrate binding to the metallonuclease domain. Interestingly, the invariant and solvent exposed R496 (R443 in PRORP2) and H498 (H445 in PRORP2) residues that reside in this region would be ideally placed for interaction with the phosphodiester backbone in the acceptor stem. We previously mutated the α21 residue H498 (H498A/H498Q, *At*PRORP1) or H445 (H445A, *At*PRORP2); all mutations reduced the STO k_{obs} without significantly affecting the K_D (Howard *et al.* 2015; Karasik *et al.* 2016). We proposed that these residues are involved in positioning the substrate and our model provides potential contacts with substrate for testing this hypothesis.

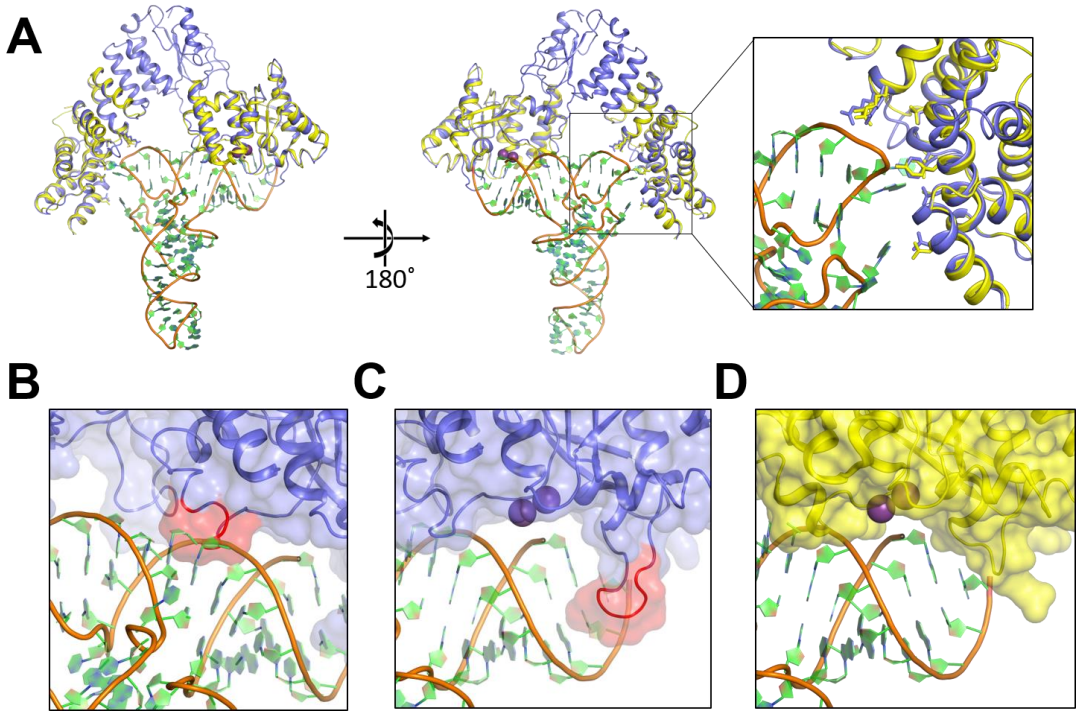


Figure 2-10 Model of the PRORP-substrate complex.

(**A**) Overall view of the modeled complex. AtPRORP2 (PDB 5diz, blue) is bound to tRNA (PDB 1ehz, orange backbone with green and blue rings). The AtPRORP1 (PDB 4g24, yellow) NYN and PPR domains are aligned to the corresponding residues domains in AtPRORP2. Close-up view of the PPR domain highlights the positions of residues for which mutation affected binding affinity (> 3-fold), including D105/N38, N136/Q70, Y140/Y74, N175/N108, T180/T113, R210/R145, and R212/R147 in AtPRORP1/AtPRORP2, respectively. (**B**) Close-up of the AtPRORP2-tRNA complex with a potential steric clash between NYN helix α21 (red) and the 3' side of the tRNA

acceptor stem. (**C–D**) Close-up of the AtPRORP2- (C) or AtPRORP1-substrate complex (D) showing the NYN active site. The loop for which AtPRORP2–3 have a 4 residue insertion is highlighted in red (C). In both structures, the 5' leader of pre-tRNA would extend forward from the panel, while the 3' trailer would extend behind the NYN domain.

Our model suggests an exit groove for the 5' leader that would place the N_{-3}/N_{-2} phosphodiester bond and N_{-3} nucleoside outside the bounds of the NYN domain. This is consistent with our Na^+-dependence data for tRNA and 1-nt pre-tRNA, which suggested few contacts beyond the N_{-2} nucleoside. Moreover, our model places the N_{-2}/N_{-1} phosphodiester bond near the invariant H438 (H386 in PRORP2) and R441 (R389 in PRORP2), implicating these residues for interactions with the negatively charged backbone. This aspect of our model is congruent with our data indicating that PRORPs interact with the tRNA leader primarily through one non-base-specific phosphodiester backbone contact. Furthermore, the α16–α17 loop is positioned to separate the 5' leader and 3' trailer. Interestingly, *At*PRORP1 has 4 fewer amino acids in the loop than *At*PRORP2/3. These loop differences might explain the variations in 5' end discrimination we observed previously, in which *At*PRORP2/3 had a stronger propensity to miscleave at the N_{-1} position when an N_{-1}/N_{73} base pair was possible (Brillante *et al.* 2016; Howard *et al.* 2016). Thus, there may be differences in the 5' end recognition employed by the two loops.

The elbow region of tRNAs are highly structured and numerous tRNA binding enzymes recognize this part of the tRNA using a variety of interaction motifs (Zhang and Ferre-D'Amare 2016). Importantly, our model predicts that: (i) R210 (R145 in *At*PRORP2) and R212 (R147 in PRORP2) would contact TψC stem phosphodiester bonds; (ii) Asp 105 (N38 in PRORP2), N136 (Q70 in PRORP2), and N175 (N108 in PRORP2) are positioned to hydrogen bond with bases of the D loop (first residue) or the TψC (last two nucleotides) loop respectively; and (iii) Y140 (Y74 in PRORP2) is capable of both base stacking and hydrogen bonding with bases of the TψC loop. Although our proposed model needs to be further tested, it provides insight into the details of

precursor tRNA binding of PRORPs in the absence of crystal structures of PRORP-tRNA complexes and a rough framework for the design of future experiments.

2.6 Conclusions

The data we present herein support a novel model for PRORP-pre-tRNA recognition that shares similarities with the mode of substrate recognition by the RNase P ribozyme. The salt-dependence of PRORP-substrate binding parses the ionic and non-ionic contributions to PRORP-substrate binding affinity. The data revealed that *At*PRORPs make at least four contacts with pre-tRNA phosphodiester bonds. Only one of these is contained in the leader sequence, most likely at the N_{-2}/N_{-1} phosphodiester bond. Additionally, we identified an extended surface on the PPR domains of *At*PRORP1 and *At*PRORP2 that interacts with substrate. Mutations on this surface suggested a mode of binding different from that of sequence specific ssRNA-binding PPR proteins. The biochemical and modeling data we have presented will facilitate the development of additional hypotheses for single-subunit PRORP substrate recognition. Given that metazoan PRORPs require two additional proteins for catalysis, there are likely differences that will need to be determined.

2.7 References

Anantharaman V, Aravind L. 2006. The NYN Domains: Novel Predicted RNAses with a PIN Domain-like Fold. *RNA Biology* **3**: 18-27.

Barkan A, Rojas M, Fujii S, Yap A, Chong YS, Bond CS, Small I. 2012. A Combinatorial Amino Acid Code for RNA Recognition by Pentatricopeptide Repeat Proteins. *PLOS Genetics* **8**: e1002910.

Barkley MD, Lewis PA, Sullivan GE. 1981. Ion effects on the lac repressor--operator equilibrium. *Biochemistry* **20**: 3842-3851.

Bonnard G, Gobert A, Arrive M, Pinker F, Salinas-Giege T, Giege P. 2016. Transfer RNA maturation in Chlamydomonas mitochondria, chloroplast and the nucleus by a single RNase P protein. *Plant J* **87**: 270-280.

Brillante N, Gossringer M, Lindenhofer D, Toth U, Rossmanith W, Hartmann RK. 2016. Substrate recognition and cleavage-site selection by a single-subunit protein-only RNase P. *Nucleic Acids Res* **44**: 2323-2336.

Chen TH, Tanimoto A, Shkriabai N, Kvaratskhelia M, Wysocki V, Gopalan V. 2016a. Use of chemical modification and mass spectrometry to identify substrate-contacting sites in proteinaceous RNase P, a tRNA processing enzyme. *Nucleic Acids Res* **44**: 5344-5355.

Chen X, Feng F, Qi W, Xu L, Yao D, Wang Q, Song R. 2016b. Dek35 encodes a PPR protein that affects cis-splicing of mitochondrial nad4 intron 1 and seed development in maize. *Mol Plant*.

Chen Y, Li X, Gegenheimer P. 1997. Ribonuclease P catalysis requires Mg2+ coordinated to the pro-RP oxygen of the scissile bond. *Biochemistry* **36**: 2425-2438.

Crary SM, Niranjanakumari S, Fierke CA. 1998. The protein component of Bacillus subtilis ribonuclease P increases catalytic efficiency by enhancing interactions with the 5' leader sequence of pre-tRNAAsp. *Biochemistry* **37**: 9409-9416.

Day-Storms JJ, Niranjanakumari S, Fierke CA. 2004. Ionic interactions between PRNA and P protein in Bacillus subtilis RNase P characterized using a magnetocapture-based assay. *RNA* **10**: 1595-1608.

deHaseth PL, Lohman TM, Record MT, Jr. 1977. Nonspecific interaction of lac repressor with DNA: an association reaction driven by counterion release. *Biochemistry* **16**: 4783-4790.

Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**: 486-501.

Fersht AR, Shi JP, Knill-Jones J, Lowe DM, Wilkinson AJ, Blow DM, Brick P, Carter P, Waye MM, Winter G. 1985. Hydrogen bonding and biological specificity analysed by protein engineering. *Nature* **314**: 235-238.

Gobert A, Gutmann B, Taschner A, Gossringer M, Holzmann J, Hartmann RK, Rossmanith W, Giege P. 2010. A single Arabidopsis organellar protein has RNase P activity. *Nature Structural & Molecular Biology* **17**: 740-U113.

Gobert A, Pinker F, Fuchsbauer O, Gutmann B, Boutin R, Roblin P, Sauter C, Giege P. 2013. Structural insights into protein-only RNase P complexed with tRNA. *Nature Communications* **4**: 1-8.

Guckian KM, Schweitzer BA, Ren RX, Sheils CJ, Tahmassebi DC, Kool ET. 2000. Factors Contributing to Aromatic Stacking in Water: Evaluation in the Context of DNA. *J Am Chem Soc* **122**: 2213-2222.

Guerrier-Takada C, Gardiner K, Marsh T, Pace NR, Altman S. 1983. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**: 849-857.

Gutmann B, Gobert A, Giege P. 2012. PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis. *Genes & Development* **26**: 1022-1027.

Hansen A, Pfeiffer T, Zuleeg T, Limmer S, Ciesiolka J, Feltens R, Hartmann RK. 2001. Exploring the minimal substrate requirements for trans-cleavage by RNase P holoenzymes from Escherichia coli and Bacillus subtilis. *Mol Microbiol* **41**: 131-143.

Hoang C, Ferre-D'Amare AR. 2002. Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell* **107**: 929-939.

Holzmann J, Frank P, Loffler E, Bennett KL, Gerner C, Rossmanith W. 2008. RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell* **135**: 462-474.

Howard MJ, Karasik A, Klemm BP, Mei C, Shanmuganathan A, Fierke CA, Koutmos M. 2016. Differential substrate recognition by isozymes of plant protein-only Ribonuclease P. *RNA* **22**: 782-792.

Howard MJ, Klemm BP, Fierke CA. 2015. Mechanistic Studies Reveal Similar Catalytic Strategies for Phosphodiester Bond Hydrolysis by Protein-only and RNA-dependent Ribonuclease P. *J Biol Chem* **290**: 13454-13464.

Howard MJ, Lim WH, Fierke CA, Koutmos M. 2012. Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5 ' processing. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 16149-16154.

Howard MJ, Liu X, Lim WH, Klemm BP, Fierke CA, Koutmos M, Engelke DR. 2013. RNase P enzymes: divergent scaffolds for a conserved biological reaction. *RNA Biol* **10**: 909-914.

Hutchison CA, 3rd, Phillips S, Edgell MH, Gillam S, Jahnke P, Smith M. 1978. Mutagenesis at a specific position in a DNA sequence. *J Biol Chem* **253**: 6551-6560.

Imai T, Nakamura T, Maeda T, Nakayama K, Gao X, Nakashima T, Kakuta Y, Kimura M. 2014. Pentatricopeptide repeat motifs in the processing enzyme PRORP1 in Arabidopsis thaliana play a crucial role in recognition of nucleotide bases at TpsiC loop in precursor tRNAs. *Biochem Biophys Res Commun* **450**: 1541-1546.

Jones S, Daley DT, Luscombe NM, Berman HM, Thornton JM. 2001. Protein-RNA interactions: a structural analysis. *Nucleic Acids Res* **29**: 943-954.

Karasik A, Shanmuganathan A, Howard MJ, Fierke CA, Koutmos M. 2016. Nuclear Protein-Only Ribonuclease P2 Structure and Biochemical Characterization Provide Insight into the Conserved Properties of tRNA 5' End Processing Enzymes. *J Mol Biol* **428**: 26-40.

Kobayashi K, Kawabata M, Hisano K, Kazama T, Matsuoka K, Sugita M, Nakamura T. 2012. Identification and characterization of the RNA binding surface of the pentatricopeptide repeat protein. *Nucleic Acids Res* **40**: 2712-2723.

Kurz JC, Fierke CA. 2002. The affinity of magnesium binding sites in the Bacillus subtilis RNase P x pre-tRNA complex is enhanced by the protein subunit. *Biochemistry* **41**: 9545-9558.

Kurz JC, Niranjanakumari S, Fierke CA. 1998. Protein component of Bacillus subtilis RNase P specifically enhances the affinity for precursor-tRNAAsp. *Biochemistry* **37**: 2393-2400.

Lai LB, Bernal-Bayard P, Mohannath G, Lai SM, Gopalan V, Vioque A. 2011. A functional RNase P protein subunit of bacterial origin in some eukaryotes. *Mol Genet Genomics* **286**: 359-369.

Latt SA, Sober HA. 1967. Protein-nucleic acid interactions. II. Oligopeptide-polyribonucleotide binding studies. *Biochemistry* **6**: 3293-3306.

Leroy JL, Gueron M, Thomas G, Favre A. 1977. Role of divalent ions in folding of tRNA. *Eur J Biochem* **74**: 567-574.

Liu X, Chen Y, Fierke CA. 2014. A real-time fluorescence polarization activity assay to screen for inhibitors of bacterial ribonuclease P. *Nucleic Acids Res* **42**: e159.
Mao G, Chen TH, Srivastava AS, Kosek D, Biswas PK, Gopalan V, Kirsebom LA. 2016. Cleavage of Model Substrates by Arabidopsis thaliana PRORP1 Reveals New Insights into Its Substrate Requirements. *PLoS ONE* **11**: e0160246.
Martin AM, Horton NC, Lusetti S, Reich NO, Perona JJ. 1999. Divalent metal dependence of site-specific DNA binding by EcoRV endonuclease. *Biochemistry* **38**: 8430-8439.
Marvin MC, Engelke DR. 2009. Broadening the mission of an RNA enzyme. *J Cell Biochem* **108**: 1244-1251.
Milligan JF, Uhlenbeck OC. 1989. Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol* **180**: 51-62.
Niranjanakumari S, Kurz JC, Fierke CA. 1998. Expression, purification and characterization of the recombinant ribonuclease P protein component from Bacillus subtilis. *Nucleic Acids Res* **26**: 3090-3096.
O'Toole N, Hattori M, Andres C, Iida K, Lurin C, Schmitz-Linneweber C, Sugita M, Small I. 2008. On the expansion of the pentatricopeptide repeat gene family in plants. *Mol Biol Evol* **25**: 1120-1128.
Pantoliano MW, Petrella EC, Kwasnoski JD, Lobanov VS, Myslik J, Graf E, Carver T, Asel E, Springer BA, Lane P et al. 2001. High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J Biomol Screen* **6**: 429-440.
Pavlova LV, Gossringer M, Weber C, Buzet A, Rossmanith W, Hartmann RK. 2012. tRNA processing by protein-only versus RNA-based RNase P: kinetic analysis reveals mechanistic differences. *ChemBioChem* **13**: 2270-2276.
Pierce BG, Wiehe K, Hwang H, Kim BH, Vreven T, Weng Z. 2014. ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics* **30**: 1771-1773.
Record MT, Jr., Anderson CF, Lohman TM. 1978. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Q Rev Biophys* **11**: 103-178.
Record MT, Jr., Lohman ML, De Haseth P. 1976. Ion effects on ligand-nucleic acid interactions. *J Mol Biol* **107**: 145-158.
Reinhard L, Sridhara S, Hallberg BM. 2015. Structure of the nuclease subunit of human mitochondrial RNase P. *Nucleic Acids Res* **43**: 5664-5672.
Reiter NJ, Osterman A, Torres-Larios A, Swinger KK, Pan T, Mondragon A. 2010. Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* **468**: 784-789.
Ringel R, Sologub M, Morozov YI, Litonin D, Cramer P, Temiakov D. 2011. Structure of human mitochondrial RNA polymerase. *Nature* **478**: 269-273.
Rueda D, Hsieh J, Day-Storms JJ, Fierke CA, Walter NG. 2005. The 5' leader of precursor tRNAAsp bound to the Bacillus subtilis RNase P holoenzyme has an extended conformation. *Biochemistry* **44**: 16130-16139.
Schmitz-Linneweber C, Small I. 2008. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci* **13**: 663-670.
Schrödinger, LLC. 2010. PyMOL Molecular Graphics System, Version 1.5.
Shi H, Moore PB. 2000. The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited. *RNA* **6**: 1091-1105.

Singh M, Wang Z, Koo BK, Patel A, Cascio D, Collins K, Feigon J. 2012. Structural basis for telomerase RNA recognition and RNP assembly by the holoenzyme La family protein p65. *Mol Cell* **47**: 16-26.

Small ID, Peeterns N. 2000. The PPR motif – a TPR-related motif prevalent in plant organellar proteins. *Trends in Biochemical Sciences* **25**: 46-47.

Sugita C, Komura Y, Tanaka K, Kometani K, Satoh H, Sugita M. 2014. Molecular characterization of three PRORP proteins in the moss Physcomitrella patens: nuclear PRORP protein is not essential for moss viability. *PLoS ONE* **9**: e108962.

Taschner A, Weber C, Buzet A, Hartmann RK, Hartig A, Rossmanith W. 2012. Nuclear RNase P of Trypanosoma brucei: a single protein in place of the multicomponent RNA-protein complex. *Cell Rep* **2**: 19-25.

Vilardo E, Nachbagauer C, Buzet A, Taschner A, Holzmann J, Rossmanith W. 2012. A subcomplex of human mitochondrial RNase P is a bifunctional methyltransferase--extensive moonlighting in mitochondrial tRNA biogenesis. *Nucleic Acids Res* **40**: 11583-11593.

Walczyk D, Gossringer M, Rossmanith W, Zatsepin TS, Oretskaya TS, Hartmann RK. 2016. Analysis of the Cleavage Mechanism by Protein-Only RNase P Using Precursor tRNA Substrates with Modifications at the Cleavage Site. *J Mol Biol* **428**: 4917-4928.

Walker SC, Engelke DR. 2006. Ribonuclease P: the evolution of an ancient RNA enzyme. *Crit Rev Biochem Mol Biol* **41**: 77-102.

Yagi Y, Hayashi S, Kobayashi K, Hirayama T, Nakamura T. 2013. Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS ONE* **8**: e57286.

Yin P, Li Q, Yan C, Liu Y, Liu J, Yu F, Wang Z, Long J, He J, Wang H-W et al. 2013. Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* **504**: 168-171.

Zhang J, Ferre-D'Amare AR. 2016. The tRNA Elbow in Structure, Recognition and Evolution. *Life (Basel)* **6**.

Chapter 3 Pentatricopeptide Repeats of Protein-only RNase P use a Distinct Mode to Recognize Conserved Bases and Structural Elements of Pre-tRNA[2]

3.1 Abstract

Pentatricopeptide repeat (PPR) motifs are α-helical structures known for their modular recognition of single-stranded RNA sequences with each motif in a tandem array binding to a single nucleotide. Protein-only RNase P 1 (PRORP1) in *Arabidopsis thaliana* is an endoribonuclease that uses its PPR domain to recognize precursor tRNAs (pre-tRNAs) as it catalyzes removal of the 5´-leader sequence from pre-tRNAs with its NYN metallonuclease domain. To gain insight into the mechanism by which PRORP1 recognizes tRNA, we determined a crystal structure of the PPR domain in complex with yeast $tRNA^{Phe}$ at 2.85 Å resolution. The PPR domain of PRORP1 bound to the structurally conserved elbow of tRNA and recognized conserved structural features of tRNAs using mechanisms that are different from the established single-stranded RNA recognition mode of PPR motifs. The PRORP1 PPR domain-$tRNA^{Phe}$ structure revealed a conformational change of the PPR domain upon tRNA binding and moreover demonstrated the need for pronounced overall flexibility in the PRORP1 enzyme conformation for

[2] The following chapter is a slightly modified version of the a publication in Nucleic Acids Research: Teramoto,T., Kaitany,K.J., Kakuta,Y., Kimura,M., Fierke,C.A. and Hall,T.M.T. (2020) Pentatricopeptide repeats of protein-only RNase P use a distinct mode to recognize conserved bases and structural elements of pre-tRNA. *Nucleic Acids Res.*, (Online only). As the second author I produced all thermodynamic fluorescent binding and single-turnover kinetic data and transcribed labeled and purified the pre-tRNA substrates used in these assays. I was a major contributor to the writing of this manuscript and played a large role in the writing of the results and conclusions of the manuscript.

substrate recognition and catalysis. The PRORP1 PPR motifs have evolved strategies for protein-tRNA interaction analogous to tRNA recognition by the RNA component of ribonucleoprotein RNase P and other catalytic RNAs, indicating convergence on a common solution for tRNA substrate recognition.

3.2 Introduction

Transfer RNA (tRNA) is an essential non-coding RNA required for protein translation that physically links the genetic code in mRNA to the amino acid sequence of a protein. With a classical length of 70-85 nucleotides, mature tRNAs fold into the functional 3D L-shaped structure. tRNAs undergo a maturation process whereby precursor molecules undergo a series of individual processing steps, including 5´- and 3´- sequence removal, RNA base modifications, splicing, and addition of the conserved 3´-terminal CCA sequence. Correct tRNA processing is critical for function. It is therefore not surprising that mutations in tRNA genes and tRNA processing enzymes are associated with multiple diseases (1).

Ribonuclease P (RNase P) is an endoribonuclease that catalyzes the 5´ maturation of tRNA precursors (pre-tRNA). Two forms exist in nature. In all three phylogenetic domains, RNase P occurs as a ribonucleoprotein (RNP) containing a catalytic RNA and up to 10 protein cofactors. In eukaryotes, protein-only RNase P (PRORP) forms are present in the nuclei and organelles (2,3). A minimal protein-only RNase P was also reported in some phyla of bacteria and archaea (4,5). Because RNP RNase P and PRORP catalyze identical reactions, they serve as a model system for how RNA-based and polypeptide enzymes have co-evolved (3). Since Altman and co-workers discovered in 1983 that *Escherichia coli* RNase P RNA (M1 RNA) is a ribozyme (6), biochemical and structural studies have focused largely on bacterial RNase P RNAs, and much information on structure-function relationships is available (2,7). The structures of bacterial, archaeal, and

eukaryotic RNP RNase P in complex with tRNA elucidated the structural basis for the catalytic mechanism and tRNA recognition of RNP RNase P at the atomic level (8-11).

How 5´ pre-tRNA processing was carried out in the apparent absence of RNP RNase P in plant cells and organelles in eukaryotic cells was a mystery until the discovery of PRORP in 2008. PRORP was initially found in human mitochondria (12,13) and subsequently found in the organelles and nuclei of the model plant *A. thaliana* (14,15), the alga *Ostreococcus tauri* (16), the protozoan *Trypanosoma brucei* (17), and the moss *Physcomitrella patens* (18). A recent bioinformatics analysis described that PRORP proteins are widely present in four out of the five main eukaryotic supergroups (apparently absent in Amoebozoa), and that the occurrence of RNP RNase P and PRORP proteins seems mutually exclusive in genetic compartments of modern Eukarya (19). Human mitochondrial PRORP was the first identified (originally termed mammalian mitochondrial RNase P, MRPP3), and it requires two additional protein subunits (MRPP1 and MRPP2) for function (12). In contrast, *A. thaliana* PRORP requires no additional subunits to catalyze pre-tRNA maturation *in vitro* (although there is some evidence that plant organellar PRORP might function with other proteins *in vivo* (20)), and as a result, it is used as a model PRORP for comparison with RNP RNase P. *A. thaliana* encodes three isozymes, PRORP1, PRORP2, and PRORP3. PRORP1 is localized in mitochondria and chloroplasts, while PRORP2 and PRORP3 are in the nuclei (14). PRORP1 can catalyze the endonucleolytic maturation of pre-tRNA and substitute for RNP RNase P activity in *E. coli* (14) as well as in *Saccharomyces cerevisiae* (21). Crystal structures of PRORP1 and PRORP2 revealed that PRORP proteins comprise five tandem pentatricopeptide repeat (PPR) motifs, a central linker domain, and a metallonuclease domain belonging to the NYN family. The N-terminal PPR domain is involved in pre-tRNA binding, while the C-terminal metallonuclease domain catalyzes cleavage (22,23).

The PRORP PPR domain has been shown to contact the D and TψC loops at the pre-tRNA 'elbow' while the NYN nuclease domain is responsible for catalysis at the 5′ end (3,22,24). No crystal structures of a PRORP in complex with pre-tRNA or tRNA have been reported, leaving open the question of how the PPR domain recognizes the structured pre-tRNA elbow.

PPR motif-containing proteins are eukaryote-specific and widely distributed among RNA-binding proteins in plants that are involved in organelle transcript processing and stability (25-27). PPR motifs are often found in tandem, and each repeat forms a helix-turn-helix fold of ~35 amino acid residues (28). Computational and biochemical analyses have identified a putative RNA recognition code for PPR proteins (29-31). Recent crystal structures of the maize chloroplast protein PPR10 and a designed PPR protein in complex with single-stranded RNA support the proposed recognition code and revealed the molecular basis for specific and modular recognition of RNA bases (32,33).

Given the established single-stranded RNA recognition mode of PPR motifs, different models have been proposed for the PRORP1 PPR domain recognition of substrate pre-tRNA. Imai *et al.* proposed a model of tRNA docked onto PRORP1 with the second (PPR2) and third (PPR3) PPR motifs recognizing nucleotides C56 and A57, respectively, in the TψC loop of pre-tRNA (34). In addition, Pinker *et al.* reported a small angle scattering (SAXS)-based model of PRORP2 in complex with pre-tRNA, showing its PPR2 and PPR3 motifs recognizing C56 in the TψC loop and G18 in the D-loop, respectively (35). Both of these models used the rules derived for sequence-specific binding of single-stranded RNA by PPR motifs and proposed that PPR2 selects a pyrimidine while PPR3 selects a purine. In contrast, Klemm *et al.* demonstrated that conserved residues outside the canonical PPR RNA-interacting positions are critical for pre-tRNA recognition and tRNA sequence substitutions had little effect on PRORP1-pre-tRNA binding

affinity (36). Moreover, the salt dependence of PRORP-pre-tRNA affinity indicated the involvement of ionic interactions. As a result, Klemm *et al.* proposed a model of the PRORP-pre-tRNA complex without sequence-specific interactions, instead showing the PPR domain interacting with backbone phosphodiester bonds of tRNA. This model is also consistent with biochemical data showing that the PPR3 motif of *A. thaliana* PRORP3 could not be 'reprogrammed' using the established single-stranded RNA-binding code (29-31) to bind to tRNA bearing a pyrimidine at nucleotide 57 (37).

To gain an atomic level understanding of how PRORP recognizes pre-tRNA and resolve the discrepancies in models of PPR motif recognition of the tRNA elbow, we determined a crystal structure of the PPR domain of *A. thaliana* PRORP1 in complex with yeast $tRNA^{Phe}$ at a resolution of 2.85 Å. The structure revealed that the PPR domain of PRORP1 recognizes the 'elbow' region of yeast $tRNA^{Phe}$ via base- and structure-specific interactions that are completely different from the established single-stranded RNA recognition by other PPR proteins. The PPR1 and PPR2 motifs recognize two invariant nucleotide residues, G19 and C56, that form a base pair between the D and TψC loops. In addition, residues within the PPR1 motif bind U17 and G20 in the D loop, and residues in the PPR3 motif interact with the TψC loop. These residues form electrostatic and stacking interactions with the tRNA and also hydrogen bond interactions that appear to be base specific. Each of these interactions contributes to the binding affinity. This tRNA elbow recognition mode is remarkably similar to that of RNP RNase P, as well as to those of 23S rRNA (38) and T-box riboswitches (39). Thus, the results presented here support the notion that PRORP proteins and functional RNAs, such as RNase P RNA, 23S rRNA and T-boxes, have converged on a similar solution to tRNA recognition (8-11,24,35,38-40). The tRNA interaction mode of the

PPR domain also reveals evolution of PPR motifs to recognize structured RNA in addition to single-stranded RNA sequences.

3.3 Materials and Methods

Protein expression and purification

The cDNA sequence encoding an engineered PPR domain (residues 84-292 with solubilizing substitutions Y266N, F284Q, and F291Q) was subcloned into the pSMT3 vector (kindly provided by Christopher Lima, Memorial Sloan Kettering Cancer Center), which encodes an N-terminal His_6-SUMO tag. The engineered PPR domain was expressed in *E. coli* strain BL21-CodonPlus (DE3)-RIL (Agilent Technologies) at 18 °C overnight after induction with 0.5 mM IPTG. The cells were collected by centrifugation, and pellets were resuspended in lysis buffer (containing 50 mM Tris-HCl, pH 8.0, 500 mM NaCl) and stored at −80 °C until use. The cells were disrupted by sonication followed by centrifugation to remove cell debris. The soluble fraction was applied to a Ni-NTA agarose column and thoroughly washed with lysis buffer containing 20 mM imidazole. The target SUMO fusion protein was eluted with lysis buffer containing 400 mM imidazole. The fusion protein was cleaved overnight with 0.2 mg of Ulp1 protease and dialyzed against a buffer containing 50 mM Tris-HCl, pH 8.0, 150 mM NaCl and 0.5 mM TCEP. The protein was then loaded onto a HiTrap Q column (GE Healthcare), which did not bind PRORP1 (Figure 3-1A). The flow-through fractions were pooled and dialyzed against a buffer containing 25 mM Hepes-NaOH, pH 7.0, 150 mM NaCl and 0.5 mM TCEP. The sample was loaded onto a HiTrap SP column (GE Healthcare). Bound proteins were eluted using a linear gradient from 0.1 to 1 M NaCl in 25 mM Hepes-NaOH, pH 7.0. Peak fractions containing the engineered PPR domain were pooled and concentrated and reducing agent was added (final concentration of 1 mM dithiothreitol or 5 mM 2-mercaptoethanol). The protein was purified further using a Superdex 75

10/300 GL column (GE Healthcare), equilibrated with 25 mM HEPES, pH 7.0, 200 mM NaCl, and 0.5 mM TCEP. Final purified protein was concentrated to 5 mg/mL.

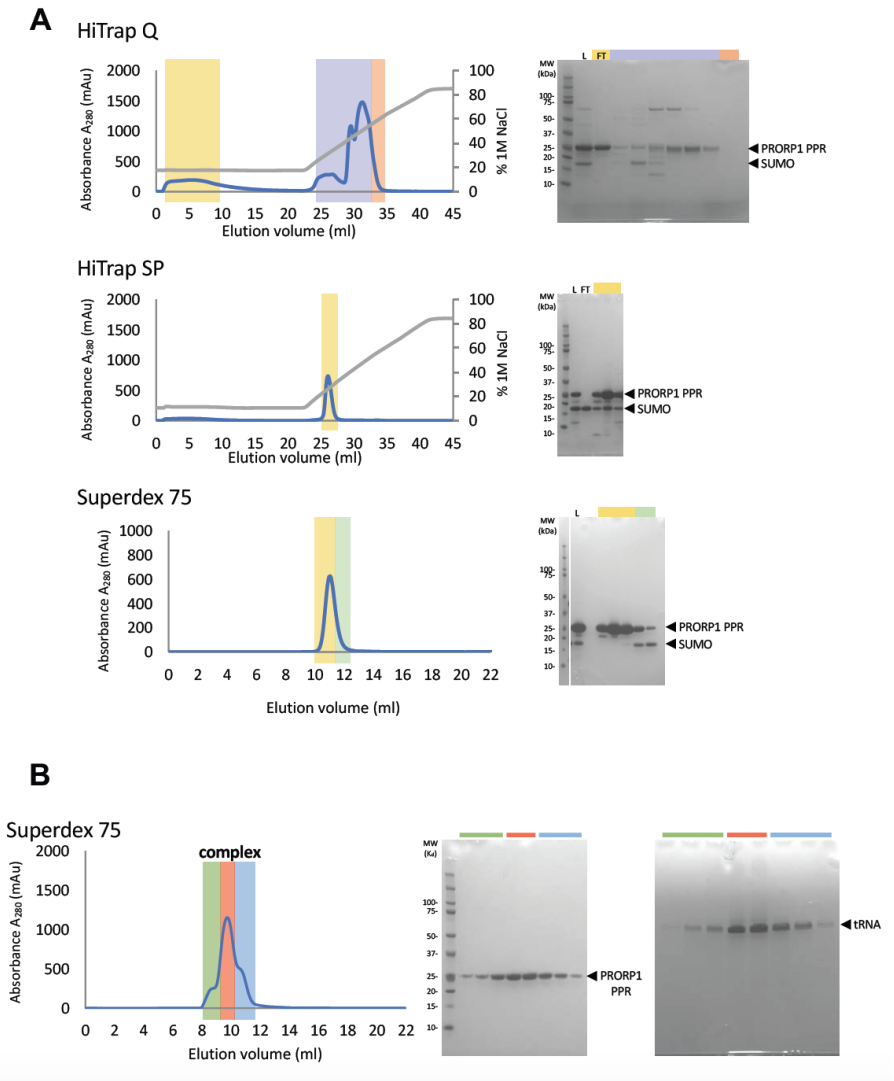


Figure 3-1 Purification of the PRORP1 PPR domain and PRORP1 PPR domain-tRNA^{Phe} complex for crystallization.

(**A**) Chromatography profiles and SDS-PAGE gels for purification of the PRORP1 PPR domain. For all chromatograms, fractions in the yellow highlighted region were pooled and loaded onto the subsequent column. For all gels, L marks the lane with the protein sample loaded onto the column and FT marks the lane with a representative flow through protein sample. The cleaved SUMO-PRORP1 PPR domain fusion protein preparation was loaded onto a HiTrap Q column to remove *E. coli* nucleic acid contamination, which bound to the column and was eluted in complex with the PRORP1 PPR domain (purple region) or as free nucleic acid (pink region). The majority of SUMO protein and PRORP1 PPR domain flowed through (yellow region). The HiTrap Q flow through fractions were pooled and loaded onto a HiTrap SP column and eluted with a 10-100% gradient of 1M NaCl. The PRORP1 PPR domain eluted within a single peak (yellow region). Much of the SUMO protein did not bind to the HiTrap SP column (see FT fraction lane). The pooled HiTrap

SP peak fractions were loaded onto a Superdex 75 column, which separated the PRORP1 PPR domain (yellow region) from residual SUMO protein (green region). (**B**) Chromatography profile, SDS-PAGE gel, and TBE-PAGE gel for purification of the PRORP1 PPR domain-$tRNA^{Phe}$ complex. The fractions in the red region were pooled and concentrated to A260 =43 for crystallization.

For crystallization, the engineered PPR domain was mixed with commercially available yeast $tRNA^{Phe}$ (Sigma-Aldrich). Protein and RNA were mixed at a ratio of 1:1.05, and the mixture was incubated at 4 °C overnight. The protein-RNA complex was purified further using a Superdex 75 10/300 GL column (GE Healthcare), equilibrated with 25 mM HEPES, pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$, and 0.5 mM TCEP (Figure 3-1B). Peak fractions containing the complex were pooled and concentrated to A_{260} = 43.

Full-length PRORP1 protein for binding assays was purified as described in a previous study (22). The protein was purified by Ni-NTA agarose chromatography, followed by purification on a HiTrap SP column. Peak fractions containing the full-length protein were pooled and concentrated. The protein was purified further using a HiLoad 16/60 Superdex 75 column (GE Healthcare) equilibrated with 20 mM MOPS, pH 7.8, 300 mM NaCl and 0.5 mM TCEP. The peak fractions were pooled and concentrated to 76 µM. Mutant proteins were expressed at equivalent levels to WT protein, behaved similarly during the purification steps, and no differences were detected in CD spectra.

Circular dichroism analyses

To assess folding of the mutant proteins, we measured circular dichroism (CD) spectra of full-length PRORP1 wild-type, R212K, R210S, and Y133Q proteins (Figure 3-2). The CD spectra were measured on a Chirascan CD spectrometer (Applied Photophysics) at 20 °C. For each sample (300 µl in a 0.1 cm light-path cell), three scans were accumulated in the wavelength range of 190–260 nm with a 0.2 nm step size. Protein samples were 0.13 mg/ml in 25 mM sodium phosphate, pH 7.5, 200 mM NaCl. The raw CD data were adjusted by subtracting a buffer blank. CD spectra of wild-type and mutant proteins displayed negative ellipticities at 208/222 nm and 215 nm, which indicate the presence of α helices and β strands, respectively.

Crystallization, data collection, structure determination and refinement

Crystallization of the purified PRORP1 PPR domain-tRNA^{Phe} complex was performed by the sitting-drop vapor diffusion method at 4 °C. Sitting drops contained 250 nL of protein-RNA complex solution mixed with 250 nL of reservoir solution (1.4-1.5 M sodium citrate). Prior to data collection, crystals were transferred to a cryoprotectant solution containing 1.6 M sodium citrate and flash cooled to -180 °C. X-ray diffraction data were collected at beamline 22-ID of the Advanced Photon Source (APS) at 100 K with a wavelength of 1.000 Å. The data were processed using HKL2000 (HKL Research Inc.) (41). Phases were determined by molecular replacement using the program Phaser and search models of the PPR domain of *Arabidopsis* PRORP1 (PDB ID: 4G24) and yeast tRNA^{Phe} (PDB ID: 1EHZ). Model building was carried out with the program Coot (42). The programs Refmac5 (43) and Phenix.refine (44) were used for refinement. The structures displayed good geometry when analyzed by MolProbity (45). Approximately 98% and 2% of the residues constituting the PPR domain were in the most favored and allowed regions of the Ramachandran plot, respectively. Modified bases were modeled into the structure: 2-methyl-

guanosine at position 10, 5,6-dihydrouridine (D) at positions 16 and 17, N2-dimethylguanosine at position 26, O2´-methyl-cytidine at position 32, O2´-methyl-guanosine at position 34, the Y base or wybutosine at position 37, pseudouridine (ψ) at positions 39 and 55, 5-methyl-cytidine at positions 40 and 49, 7-methyl-guanosine at position 46, 5-methyl-uridine at position 54, and 1-methyl-adenosine at position 58. The average B factor for the tRNA is high due to poor electron density in regions of the tRNA that do not contact the PRORP1 protein. However, the electron density is strong at the tRNA D and TψC loops where it contacts PRORP1 (Figure 3-3A).

***In vitro* transcription**

Pre-tRNAs were synthesized as previously described (36,46) through run-off transcription from a restriction-digested (BstNI) pUC18 plasmid encoding *Bacillus subtilis* pre-tRNA^{Asp}. *In vitro* transcription reactions were run in 5:1 excess of 5´-O-monophosphorothiate guanosine (GMPS) to GTP. The 5´-GMPS pre-tRNA product was reacted with 5-iodoaceamidofluorescein (5-IAF) in a 1:40 molar ratio (RNA:5-IAF) to produce a 5´-fluorescein labeled product. Labeling reactions were carried out in 10 mM Tris, pH 7.2 with 1 mM EDTA for 16 h at 37 °C yielding 25-30% fluorescently-labeled pre-tRNA. The labeled pre-tRNA was gel purified using 12% urea-PAGE and eluted using the crush-soak method (47). Purified products were concentrated using 10 kDa MWCO Amicon® Ultra Centrifugal Filters, and ethanol precipitation. Pre-tRNA stocks were resuspended in 10 mM Tris, pH 8.0 with 1 mM EDTA, quantified by absorbance, and stored at -80 °C.

The extinction coefficient for the *B. subtilis* pre-tRNA^{Asp} at 260 nM was experimentally determined to be 674,390 M^{-1} cm^{-1} through alkaline hydrolysis. Concentrations of fluorescein were measured at 492 nm (extinction coefficient = 78000 M^{-1} cm^{-1}). Prior to all assays, substrates were

thawed, diluted with nuclease-free water, and heated at 95 °C for 90 sec followed by refolding by incubating at 25 °C for ≥ 15 min, and then incubating with metal-containing buffer for ≥ 15 min.

Fluorescence anisotropy binding assays

Binding assays were performed in Corning black polystyrene half-area 96-well plates (Product number 3686), as previously described (22,46). In short, PRORP1 variants were serially diluted from 20 μM to 9 nM and equal volumes of enzyme and 20 nM 5´-fluorescein-pre-tRNA substrate were mixed; a minimum of 12 concentrations was analyzed. Enzyme-substrate mixtures were incubated at 25 ± 1 °C in 30 mM MOPS, pH 7.8, 330 mM NaCl, 1 mM TCEP, and 20 mM $CaCl_2$. Anisotropy readings of the 5´-fluorescein-pre-tRNA tag were measured with a Tecan Ultra plate reader using an excitation wavelength of 485 nm, and emission wavelength of 535 nm. The anisotropy measurements at each enzyme concentration were observed 5 times over 15-20 min to ensure complete equilibration. The concentration dependence of the anisotropy changes was well described by a single binding isotherm (Equation 1) (where *FA* is the fluorescence anisotropy, FA_0 is the initial anisotropy, Δ*FA* is the total change in anisotropy, P is the concentration of PRORP and K_D is the dissociation constant). The K_D values and standard error for K_D values were calculated by fitting Eq. 1 to the data points from a single experimental trial using GraphPad Prism to carry out non-linear regression analysis.

Equation (1) $$FA = FA_0 + \frac{\Delta FA \cdot [P]}{[P] + K_D}$$

Single-turnover kinetic assays

Single-turnover assay reactions were initiated through addition of 5–45 μM enzyme to an equal volume of 30 nM 5´-fluorescein-pre-tRNA substrate. Reactions were carried out at 25 ± 1 °C in 30 mM MOPS, pH 7.8, 330 mM NaCl, 1 mM TCEP, and 20 mM $MgCl_2$. At specific time

points (0–4800 sec), 4 µL aliquots of the reaction were quenched with an equal volume of 100 mM EDTA, 8 M Urea, 0.05% bromophenol blue, 0.05% xylene cyanol, and 2 µg/µL yeast tRNA. Fluorescently labeled pre-tRNA substrate and 5′ leader product were separated by electrophoresis on 22.5% denaturing urea-PAGE gel. Gels were visualized using an Amersham Typhoon Biomolecular Imager, and the fraction of product was quantified using ImageJ software. A minimum of 10 time points was analyzed for each mutant. Observed single-turnover rate constants and standard errors were obtained by fitting a single exponential to the data points from a single experimental trial using GraphPad Prism 8 (Equation 2).

Equation (2) $$Fraction\ product = A - B(e^{-kt})$$

Electrophoretic mobility shift assays

tRNAs were radiolabeled at the 5′ end using [γ-^{32}P] ATP and T4 polynucleotide kinase, then purified using an Illustra MicroSpin G-25 column (GE Healthcare). RNA-binding reactions included 0.9 nM radiolabeled RNA and protein serially diluted (2-fold) from 25 µM to 3 nM. Binding reactions were incubated for 1 h at 20 °C in 20 mM HEPES, pH 7.9, 100 mM NaCl, 1 mM TCEP, 10 mM $CaCl_2$, 0.02% (v/v) Tween 20, 0.1 mg/mL poly r(U) , and 2.5% (v/v) glycerol and separated by electrophoresis on 10% TBE polyacrylamide gels (Invitrogen). Gels were dried and exposed to storage phosphor screens for 6–20 h, scanned with a Typhoon 8600 Imager, and the band intensities were quantified with ImageQuant 5.2. The data for three technical replicates were analyzed and K_D values were calculated via non-linear regression analysis for one-site binding with GraphPad Prism 7.

3.4 Results

Engineering a PRORP1 PPR domain for crystallization

To understand substrate recognition by PRORP enzymes, we sought to determine a crystal structure of a PRORP in complex with tRNA. Through protein engineering, we obtained crystals suitable for structure determination of the PPR domain of *A. thaliana* PRORP1 in complex with $tRNA^{Phe}$. Our attempts to crystallize a PRORP containing both PPR and catalytic domains in complex with tRNA were unsuccessful. We therefore focused on the PRORP1 PPR domain as the module that drives tRNA recognition (24,34-36). We engineered three regions of the PPR domain to promote crystallization. First, we noted that in the crystal structure of *Arabidopsis* PRORP1 (residues 76-572, PDB ID: 4G24) helices α10 and α11 of the PPR domain form a hydrophobic interface with the central zinc-finger domain (22). Three aromatic residues within the PPR domain (Tyr266, Phe284 and Phe291) are located at the interface. To increase the solubility of the PPR domain, we substituted these aromatic residues with hydrophilic residues (Y266N, F284Q and F291Q). Second, previous studies demonstrated that the N-terminal flexible region is involved in tRNA binding and lysine residues in this region might contact tRNA (34,40), although residues 76-94 were disordered in the crystal structure of PRORP1. We defined the minimal N-terminal region required for tRNA binding by measuring binding of N-terminal deletions of the PPR domain (named N76, N83 or N86 to indicate the N-terminal residue with all fragments extending to the C-terminal residue 294) to yeast $tRNA^{Phe}$ by electrophoretic mobility shift assay (EMSA). The affinity of the N86 variant for tRNA was considerably weaker than that of the N76 and N83 variants (Table 3-1), indicating that the N83 variant is optimal for tRNA binding affinity. Third, a long loop (LAEAATESSP) between the two α-helices of the PPR2 motif is shorter in the other *Arabidopsis* isoforms, PRORP2 and PRORP3. Substituting this loop with a shorter loop

(LASASS) had little effect on tRNA binding affinity (Table 1, PRORP1 PPR N83 ΔPPR2 loop). The engineered PRORP1 PPR domain (84-294) with three solubilizing substitutions (Y266N, F284Q and F291Q) and a shorter loop in the PPR2 motif produced well behaved protein that retained tRNA recognition and was used successfully for crystallization in complex with yeast $tRNA^{Phe}$. For simplicity, we will refer to this as the PRORP1 PPR domain.

PRORP1 PPR Domain Variant	**tRNA**	K_D **(nM)**[a]
PRORP1 PPR N76	Yeast $tRNA^{Phe}$	627 ± 55
PRORP1 PPR N83	Yeast $tRNA^{Phe}$	281 ± 44
PRORP1 PPR N83 ΔPPR2 loop	Yeast $tRNA^{Phe}$	317 ± 92
PRORP1 PPR N86	Yeast $tRNA^{Phe}$	3405 ± 338
PRORP1 PPR N83	*Arabidopsis* mito pre-$tRNA^{Cys}$	257 ± 54

Table 3-1 Binding affinities of PRORP1 PPR domain variants

[a] Measured using electrophoretic mobility shift assays in 20 mM HEPES, pH 7.9, 100 mM NaCl, 1 mM TCEP, 10 mM $CaCl_2$, 0.02% (v/v) Tween 20, 0.1 mg/mL poly r(U) , and 2.5% (v/v) glycerol at room temperature. The mean and standard error of the mean values for three technical replicates are reported.

Structure description

We determined a crystal structure of the *Arabidopsis* PRORP1 PPR domain in complex with yeast $tRNA^{Phe}$ at 2.85 Å resolution by molecular replacement (Table 3-2). Two independent crystallographic complexes were present in an asymmetric unit. The N-terminal 11 residues of the PPR domain (SRKAKKKAIQQ) are disordered in the crystal structure in complex with $tRNA^{Phe}$, despite their importance for tRNA binding (Table 3-1). The basic residues of the N-terminal flexible region could interact non-specifically with negatively-charged phosphate groups of the tRNA backbone to enhance substrate binding affinity. The two complexes include chain A (PPR domain)-chain B ($tRNA^{Phe}$) and chain C (PPR domain)-chain D ($tRNA^{Phe}$) and are highly similar

overall (root mean square deviation [rmsd] value of 1.00 Å over 192 CA atoms in the protein, 1.36 Å over 1569 atoms in the tRNA, and 1.94 Å over 1352 main chain atoms in the complex). However, each PPR domain in the asymmetric unit binds its corresponding $tRNA^{Phe}$ in a slightly different manner. The chain D tRNA molecule appears to be influenced by crystal packing forces, resulting in the chain C-chain D complex lacking some interactions. Hence, we focus on the chain A-chain B complex to describe the PPR domain-tRNA interaction.

Table 3-2 Crystallographic Summary of PRORP1 PPR-$tRNA^{Phe}$

Data Collection	
Space group	$P2_12_12$
Cell Dimensions	
a, b, c (Å)	84.5, 131.7, 155.6
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)	50.0-2.85 (2.90–2.85)*
No. unique reflections	42492
R-pim	0.055 (0.48)*
R-means	0.124 (1.08)*
CC1/2	(0.79)*
I/σI	13.8 (1.6)*
Completeness (%)	99.9 (100.0)*
Redundancy	5.0 (4.8)*

*Values in parentheses are for the highest resolution shell.

Refinement	
Resolution (Å)	49.3–2.85
R_{work} /R_{free} (%)	23.5/25.5
No. of Atoms	
Protein	3055
RNA	3136
Water	35
B-Factors	
Protein	77.1
RNA	200.4
Water	67.8
RMS Deviations	
Bond lengths (Å)	0.02
Bond angles (°)	0.6
Ramachandran plot (%)	
Favored	2.1
Allowed	97.9
PDB ID	6LVR

The crystal structure of the PRORP1 PPR domain-tRNA complex revealed that the PPR domain undergoes conformational changes that place PPR motifs 1-4 in position to interact with the tRNA. The PRORP1 PPR domain comprises 5 consecutive PPR repeats and one additional C-terminal helix (Figure 3-4A). The PPR5 motif does not interact with the tRNA. Instead it may aid in positioning the PPR1-4 motifs for tRNA elbow recognition relative to the nuclease active site. As noted in the crystal structure of full-length PRORP1, the central linker domain interacts with

the PPR5 motif and the terminal α-helix of the PPR domain (22). PPR5 together with the central linker domain bridges the tRNA elbow recognition and catalytic domains. It also serves as a C-terminal cap to the PPR domain, which stabilizes the terminal α-helices in the PPR4 motif (48). When compared with the PPR domain in the structure of full-length apo PRORP1 (PDB ID: 4G24), the first three PPR repeats (PPR1, PPR2 and PPR3) are different in their configurational details. The $tRNA^{Phe}$-bound PPR domain exhibits a more curved conformation (Figure 3-3B). With the PPR4 and PPR5 motifs aligned, the PPR3 motif in the complex is shifted away from the tRNA, whereas PPR1 is closer to the tRNA molecule. These changes allow PPRs 1-4 to interact with the tRNA, inducing a more extensive interaction surface than had been predicted. Conformational flexibility is a common feature of PPR proteins. Previous studies show that PPR proteins utilize considerable structural adaptability to bind to single-stranded RNA (32,49). In contrast, the overall structure of yeast $tRNA^{Phe}$ is unaltered by the binding of the PPR domain. The structure of yeast $tRNA^{Phe}$ is highly similar to the previously determined structure of the tRNA alone (PDB ID: 1EHZ, rmsd value of 1.56 Å over 1568 atoms).

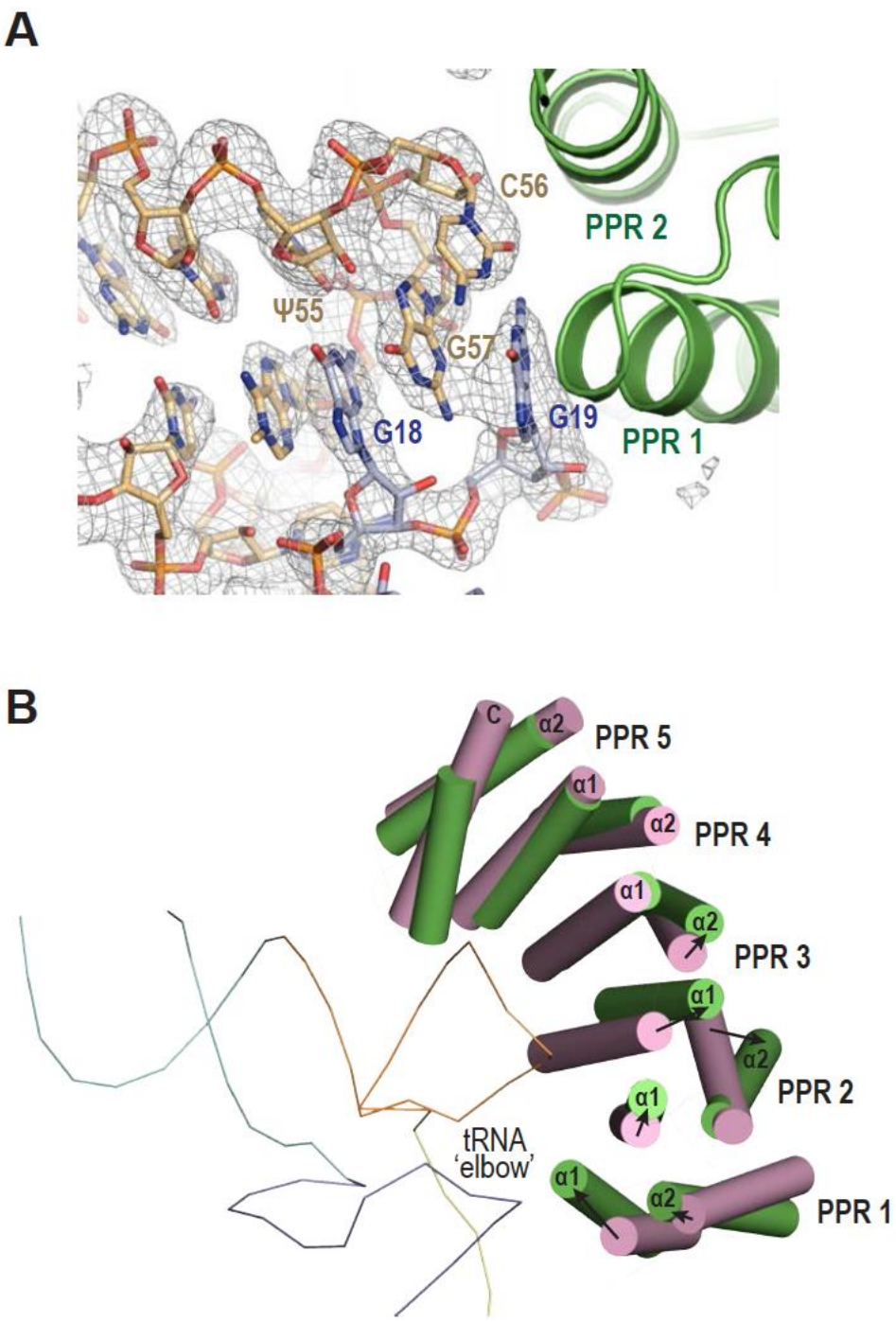


Figure 3-2 The PRORP1 PPR domain changes conformation to recognize the tRNA 'elbow'.

(**A**) The tRNA elbow binds to PRORP1 near PPR motifs 1 and 2. An Fo-Fc omit electron density map contoured at 3.0σ is superimposed with the tRNA. (**B**) Superposition of the PPR domains of the tRNA-bound PRORP1 PPR domain (green) and tRNA-free PPR domain (pink, PDB ID: 4G24). PPR domains are shown with α helices as cylinders, and the tRNA is shown as a backbone trace. Shifts in conformation are highlighted by arrows.

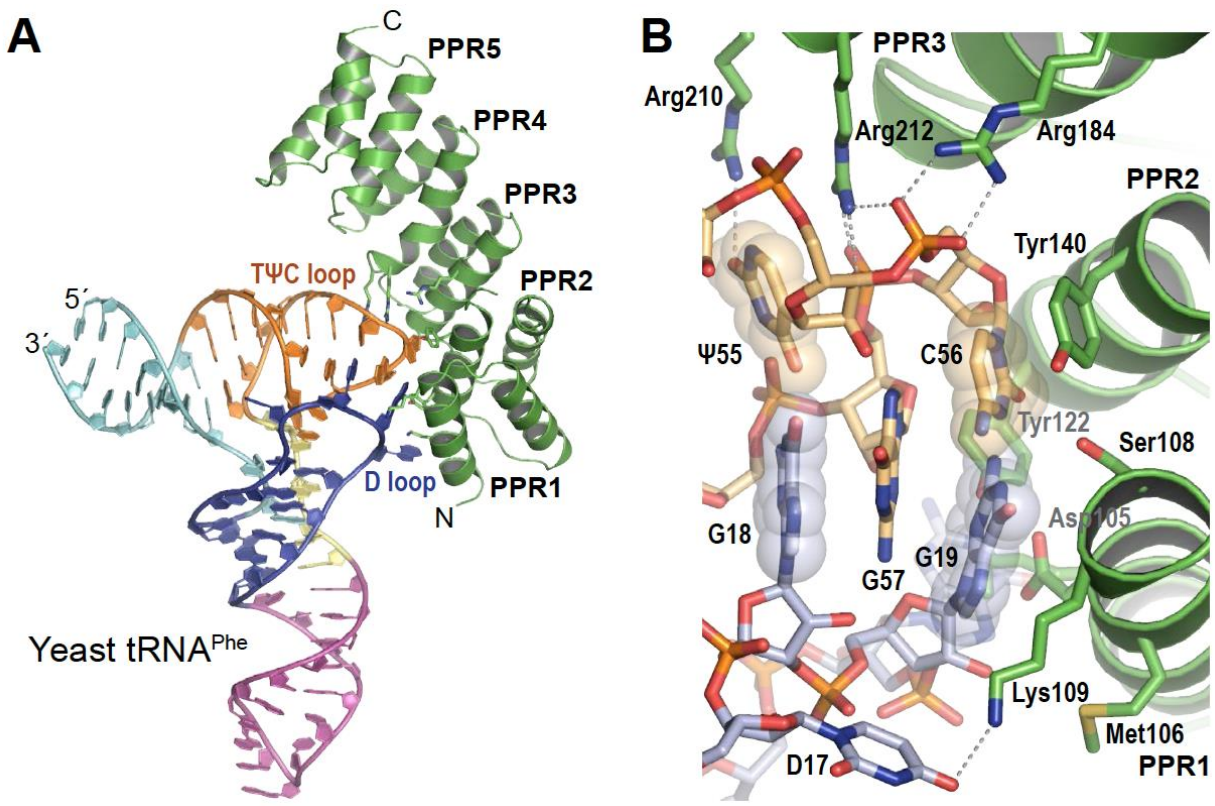


Figure 3-3 The PRORP1 PPR domain recognizes the tRNA D and TψC loops.

(**A**) Ribbon diagram of the crystal structure of the PRORP1 PPR-tRNA^{Phe} complex. PRORP1 PPR is shown in green with tRNA-interacting residues displayed as stick models. The tRNA is shown as a cartoon colored by region: acceptor stem (cyan), D stem loop (blue), anticodon stem loop (magenta), variable region (yellow), and TψC stem loop (orange). (**B**) Close-up view of the G19-C56 base pair accommodation pocket of the PRORP1 PPR domain. PRORP1 and tRNA are colored as in (A) with atom colors: oxygen (red), nitrogen (blue), phosphorus (orange), and sulfur (yellow). Dashed lines indicate interactions between PRORP1 and tRNA, and the G19-C56 and G18-ψ55 base pairs are indicated by transparent spheres.

The PRORP1 PPR domain nestles the tRNA elbow in a pocket formed by PPRs 1-4

The *Arabidopsis* PRORP1 PPR domain recognizes conserved features of the 'elbow' of the L-shaped tRNA, formed by the D and TψC loops (Figure 3-4). A ubiquitous G19-C56 tertiary interaction between the D and TψC loops is located at the tip of the tRNA elbow, and the PPR domain forms a pocket that accommodates the G19-C56 base pair (Figure 3-4B). Residues that contact the TψC loop are more conserved than those that contact the D loop (Figure 3-5). The G19

base is surrounded by residues from the PPR1 motif (Asp105, Met106, Ser108, and Lys109) and Tyr133 from the PPR2 motif (Figures 3-4B and 3-6). The OH-group of Tyr133 is hydrogen bonded to the N2 atom of the G19 base, potentially a base-specific interaction (Figure 3-6). The C56 base forms a stacking interaction with the phenol ring of Tyr140 in the PPR2 motif (Figures 3-4B and 3-6). Together these interactions appear to recognize the structure as well as the sequence of the conserved base pair. In addition to recognizing the G19-C56 base pair, PRORP1 interacts with the base pair between G18 (D loop) and the pseudouridine, ψ55 (TψC loop). The $tRNA^{Phe}$ in our structure was obtained from yeast, so it has 14 post-transcriptional modification sites (50). The guanidinium group of Arg210 in the PPR3 motif is hydrogen bonded to the O2 atom of the ψ55 base (Figures 3-4B and 3-6). The intercalation of G57 between these two base pairs forms the tRNA elbow's structural core whose sequence and structure are probed by PRORP1.

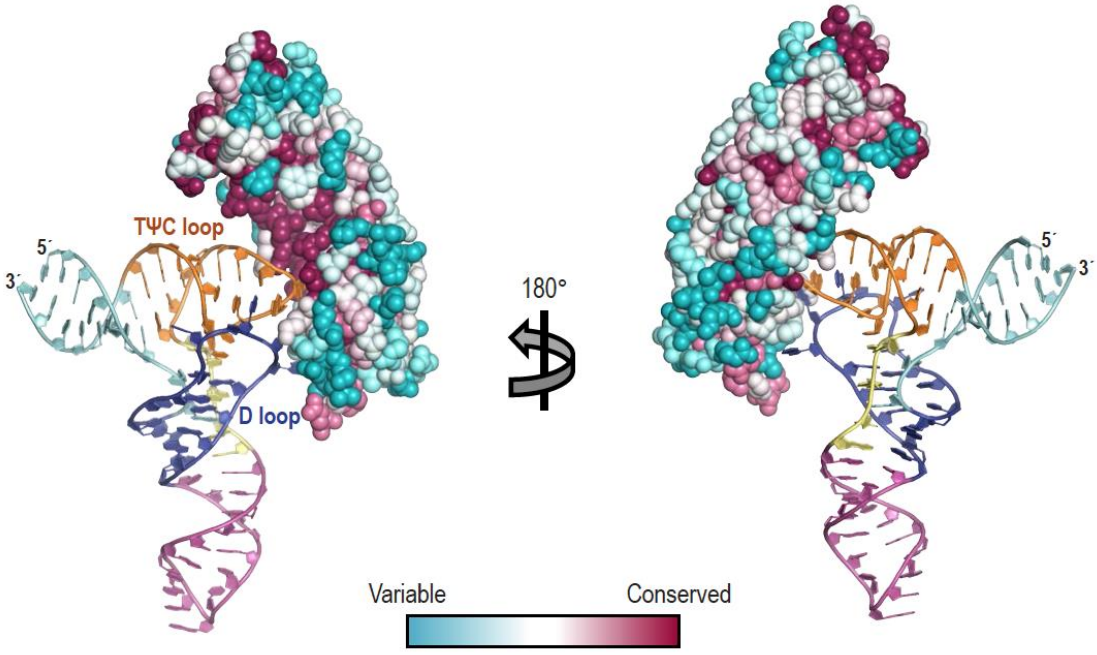


Figure 3-4 PRORP1 uses a conserved surface to interact with the TψC loop of tRNA.

The PRORP1 PPR domain is shown as a space-filling sphere representation with residues colored by degree of conservation. The ConSurf server was used to identify sequence homologs and project the degree of conservation on the structure of the PRORP1 PPR domain. The tRNA is shown as a cartoon colored by region: acceptor stem (cyan), D stem loop (blue), anticodon stem loop (magenta), variable region (yellow), and TψC stem loop (orange).

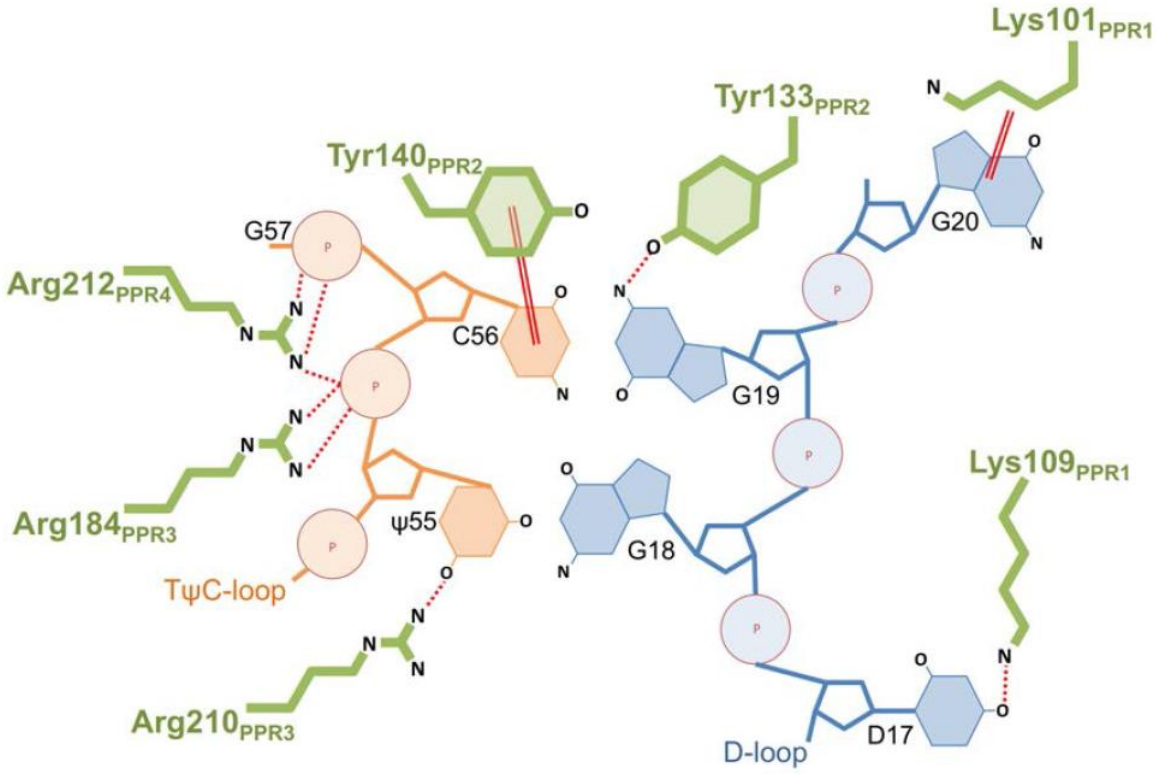


Figure 3-5 PRORP1 PPR-tRNA interactions.

Schematic representation of interactions between the PRORP1 PPR domain and yeast $tRNA^{Phe}$. PRORP1 PPR residues are green, tRNA TψC-loop nucleotides are orange, and tRNA D-loop nucleotides are light blue. Circles represent tRNA phosphate groups (P). Dotted and double lines indicate hydrophilic and stacking interactions, respectively.

The PRORP1 PPR domain forms a variety of additional tRNA interactions using basic side chains (Figure 3-6). Two unpaired nucleotides in the D loop are recognized by lysine side chains. Lys101 in the PPR1 motif stacks with the G20 base (Figure 3-6). The ε-amino group of Lys109 in the PPR1 motif makes a hydrogen bond with the O4 atom of the 5,6-dihydrouridine (D) base, D17 (Figures 3-4B and 3-6). This appears to be a base-specific contact recognizing a modified nucleotide but it is also capable of recognizing the O4 atom of an unmodified uracil. The phosphate backbone of the TψC loop is contacted by two arginine residues. The guanidino groups of Arg184 in the PPR3 motif and Arg212 in the PPR4 motif make salt-bridge interactions with the phosphate

moieties of C56 and G57, respectively (Figures 3-4B and 3-6). These interactions are consistent with the salt dependence of PRORP-pre-tRNA affinity, which suggested up to three direct contacts with substrate backbone phosphodiester bonds by the PPR domain (36).

As described above, we observed PPR domain base-specific interactions with two modified bases in tRNA, D17 and ψ55. To gain insight into the importance of base modifications for tRNA affinity, we used EMSAs to compare the K_D values for binding of the isolated PPR domain to *in vitro* transcribed *Arabidopsis* mitochondrial pre-tRNA^{Cys} lacking modified nucleotides to that of modified tRNA^{Phe}. These K_D values of 281 ± 44 nM for yeast tRNA^{Phe} and 257 ± 54 nM for *Arabidopsis* mitochondrial pre-tRNA^{Cys} are comparable (Table 3-1), suggesting that the post-transcriptional modifications are dispensable for tRNA recognition by the *Arabidopsis* PRORP1 PPR domain. The contacts between Lys109-D17 and Arg210- ψ55 that we observed in the crystal structure could be substituted by interaction of the side chains with the O4 atom of an unmodified uracil.

Our crystal structure illustrating that the PRORP1 PPR domain specifically binds the tRNA elbow fully explains previous mutagenesis experiments that identified amino acids residues in the PRORP1 PPR domain that are critical for tRNA binding (36,40). Klemm *et al*. generated a number of PRORP1 mutations and analyzed the binding affinity of full-length PRORP1 to *B. subtilis* pre-tRNA^{Asp} (36). This analysis identified Tyr133, Tyr140, Arg184 and Arg212 as essential residues for tRNA binding (Table 3-4). Mutation of each of these residues to alanine substantially reduced binding affinity. Our crystal structure is in agreement with these results, as each of these amino acid residues is located at the interface between the PPR domain and tRNA^{Phe} (Figures 3-4B and 3-6). Furthermore, it appears that the interaction of Arg212 requires specific contacts that cannot be satisfied by a lysine at this position; we found that the R212K mutation in PRORP1 also

dramatically reduced binding affinity of pre-tRNA^{Asp}, as measured by fluorescence anisotropy (Table 3-3). The purified R212K mutant protein expressed similarly to WT protein and maintains protein structure as determined by CD analysis. Chen *et al.* combined chemical modification of lysines with multiple-reaction monitoring mass spectrometry and identified Lys101 and Lys109 as putative tRNA-contacting residues (40). They further showed mutation of these lysine residues to alanine had a small to moderate effect on pre-tRNA binding by full-length PRORP1 (Table 3-4). Our crystal structure also agrees with this finding, as these two residues recognize D-loop nucleotides through stacking and hydrogen bond interactions, respectively (Figure 3-6). The consistency between our crystal structure of the PRORP1 PPR domain bound to a tRNA and activity assays with mutant full-length proteins indicates that the structure accurately depicts interactions that are critical for enzyme substrate recognition.

PRORP1 PPR motifs have evolved to recognize conserved tRNA bases and structural features

Our crystal structure confirms that the PRORP1 PPR motifs recognize tRNA using different mechanisms than observed for single-stranded RNA recognition. PPR motifs are often used for single-stranded RNA binding and recognize specific RNA bases modularly, using amino acid side chains at positions 2, 5 and 35 (alternatively 1, 4, and ii) of the PPR motif (29,30). Based on the RNA recognition code of PPR motifs, it was predicted that Asn136 and Asn175 of the PPR2 motif of PRORP1 would recognize a pyrimidine, and Thr180 and Arg210 of PPR3 would recognize a purine (34,35). Contrary to this prediction, our crystal structure reveals that Asn136, Asn175, and Thr180 do not contact the tRNA (Figure 3A, B). Residues at position 35 of PRORP1 PPR3 (Arg210) and position 2 of PPR4 (Arg212) do interact with the ψ55 base and RNA backbone, but in a completely different manner than RNA base recognition by other PPR proteins

(Figure 3-7A, B). Arg210 at position 35 of PPR3 recognizes the ψ55 base and Arg212 at position 2 of PPR4 interacts with backbone phosphate groups (Figure 3-7A). In contrast, residues at positions 2, 5, and 35 of repeat 2 of PPR10 recognize an uracil base using stacking and hydrogen bond interactions (Figure 3-7B). Y133 at position 2 of PRORP1 PPR2 also interacts with the tRNA, but it is turned away from the prospective PPR base recognition pocket to hydrogen bond with G19 (Figure 3-7C).

Table 3-3 Binding affinities and kinetic constants of full-length PRORP1 to 5´-fluorescein-pre- tRNA^{Asp}

PRORP1 Variant	k_{obs} (min^{-1})[a]	k_{obs} relative to WT	K_D (nM)[b]	K_D relative to WT	$\Delta\Delta G^0_{binding}$ (kcal/mol)
Δ76 WT	2.62 ± 0.28	1	390 ± 30	1	0
Y133D	1.17 ± 0.22	0.45	11100 ± 1600	28.4	1.9
Y133Q	2.76 ± 0.34	1.05	3200 ± 600	8.2	1.2
R210A	1.59 ± 0.20	0.61	4200 ± 400	10.8	1.4
R210S	0.49 ± 0.09	0.19	10600 ± 1000	27.2	1.9
R212K	n.d.[c]	n.d.	≥ 50000[d]	≥ 130	≥ 2.9

[a] Measured using single turnover cleavage assays in 30 mM MOPS, pH 7.8, 330 mM NaCl, 1 mM TCEP, and 20 mM $CaCl_2$ at 25 ± 1 °C. The k_{obs} and standard error values are determined from nonlinear least squares regression of a single exponential equation to the data points from a single experimental trial using GraphPad Prism.

[b] Measured using fluorescence anisotropy in 30 mM MOPS, pH 7.8, 330 mM NaCl, 1 mM TCEP, and 20 mM $CaCl_2$ at 25 ± 1 °C. The K_D and standard error values are determined from non-linear least squares regression of a single binding isotherm to the data points from a single experimental trial using GraphPad Prism.

[c] Based on the binding affinity data we were unable to measure single turnover kinetics for the R212K mutant under saturating enzyme concentration.

[d] The lower limit for the K_D for R212K was estimated from the observation of <10% change in anisotropy at 20 μM R212K mutant, assuming that the total change in anisotropy upon binding Fl-pre-tRNA was similar to that of the other PRORP1 variants.

Table 3-4 Summary of effects on binding affinity by mutation of PRORP1 tRNA-interacting residues

PRORP1 Variant	K_D (nM)	K_D relative to WT
Δ76 WT	390 ± 30[a]	1
WT (Chen)	250 ± 34[b]	1
Δ76 WT (Klemm)	155 ± 20[c]	1
K101A	859 ±159[b]	3.4
K109A	389 ± 35[b]	1.6
Y133A	2600 ± 200[c]	17
Y133F	4600 ± 300[c]	30
Y133D	11100 ± 1600[a]	28
Y133Q	3200 ± 600[a]	8
Y140A	29700 ± 7000[c]	192
Y140F	1000 ± 200[c]	6.5
R184A	10400 ± 2400[c]	67
R184K	1900 ± 300[c]	12
R210A	4200 ± 400	11
R210S	10600 ± 1000[a]	27
R212A	>30,000[c]	>194
R212K	>50000[a]	>130

[a]This work (unmodified *B. subtilis* pre-tRNA^{Asp}, Δ76 PRORP1 full-length, 30 mM MOPS (pH 7.8), 330 mM NaCl, 1 mM TCEP, 20 mM $CaCl_2$)

[b]Chen, et al. (40) (3'FTSC-labeled pre-tRNA^{Cys} with 5-nt 5' leader and 23-nt 3' trailer, PRORP1, 20 mM HEPES (pH 7.2), 10 mM $Ca(OAC)_2$, 100 mM NH4OAc, 4 mM DTT, 5% glycerol

[c]Klemm, et al. (36) (unmodified *B. subtilis* pre-tRNA^{Asp}, Δ76 PRORP1 full-length, 30 mM MOPS (pH 7.8), 330 mM NaCl, 1 mM TCEP, 20 mM $CaCl_2$)

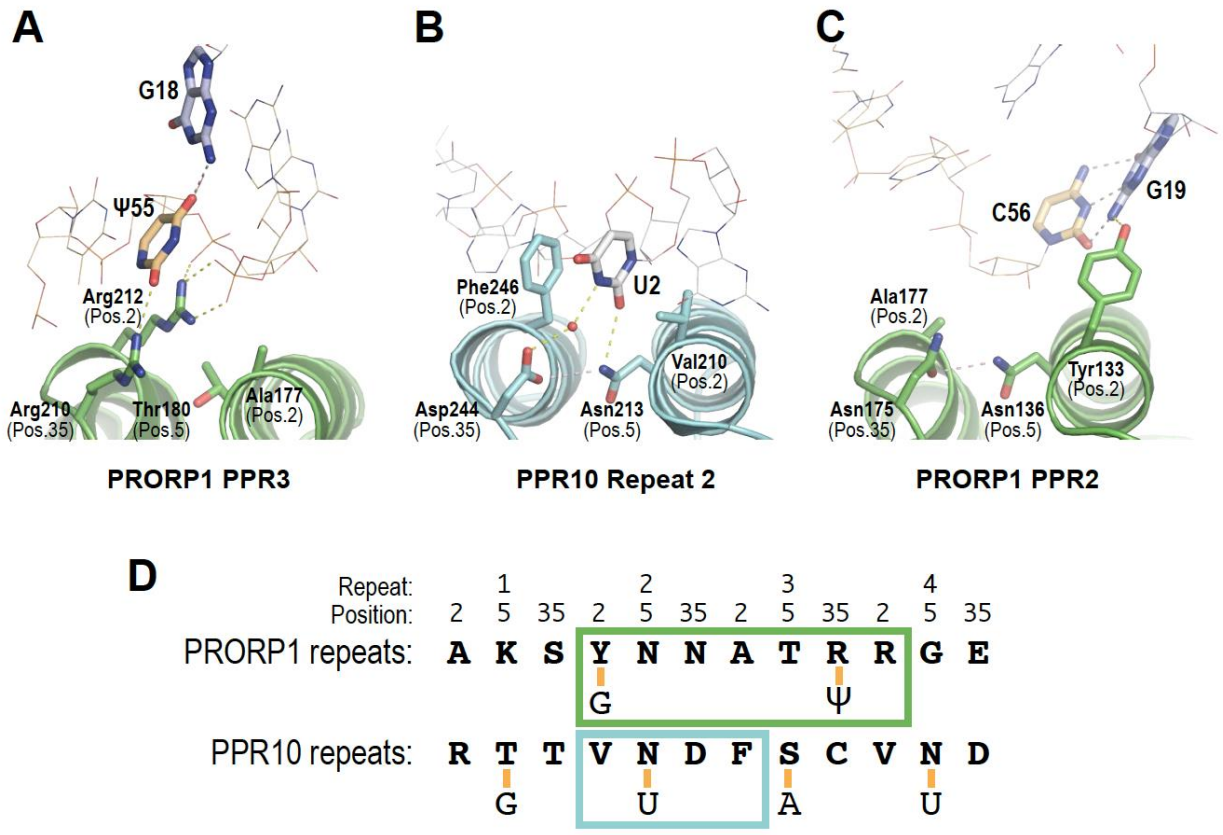


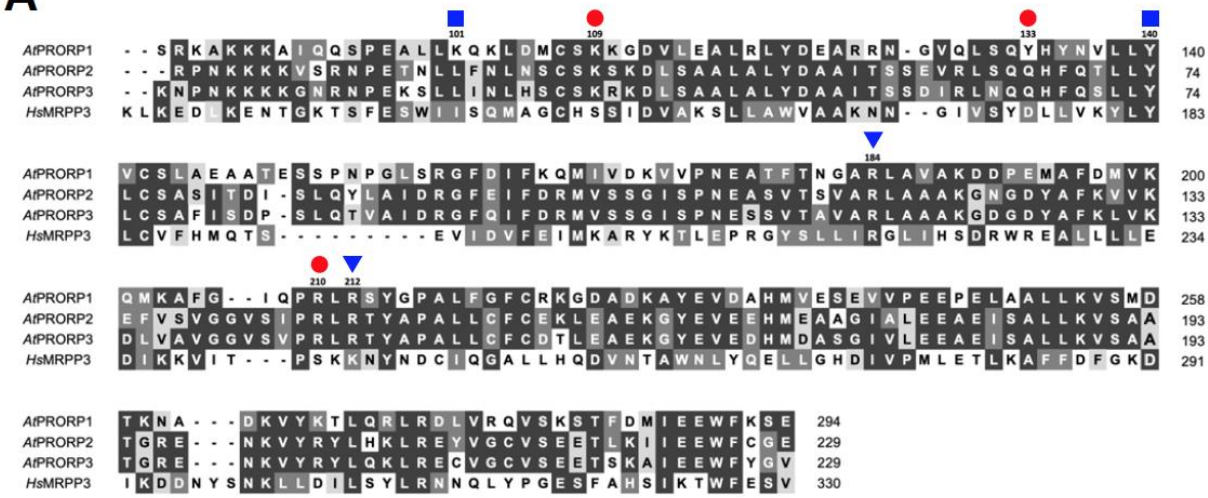
Figure 3-6 PRORP1 PPR motifs use distinct mechanisms for tRNA recognition.

(**A**) Recognition of ψ55 by PRORP1 repeat 3. (**B**) Recognition of U2 by PPR10 repeat 2 (PDB ID: 4M59). (**C**) Recognition of G19 by PRORP1 repeat 2 (PDB ID: 6LVR). (**D**) Differences in specific RNA recognition by PRORP1 and PPR10. RNA-interacting residues in PPR motifs 1-4 of PRORP1 and PPR10 are shown. Nucleotides recognized are shown below and connected by a yellow line. Individual recognition modules are boxed with PRORP1 colored green and PPR10 colored blue.

We identified several PRORP1-tRNA contacts that suggest base-specific recognition and found that Tyr133 and Arg210 are critically important for PRORP1 binding affinity and catalytic activity. No tRNA sequence substitution has yet been shown to affect PRORP recognition, but prior experiments to look for base recognition were based on the single-stranded RNA PPR recognition code and the corresponding premise that PRORP1 PPR2 interacts with a pyrimidine in the tRNA. In our crystal structure, Tyr133 interacts with G19 and together with Tyr140 appears

to recognize the G19-C56 base pair (Figure 3-6). Binding by Tyr133 and Tyr140 assures selection of this conserved feature. Mutation of Tyr133 to Phe strongly affects tRNA recognition (Table -3-4) (36), which suggests the importance of the interaction between the Tyr133 OH and G19 N2 groups. Many of the tRNA-contacting residues identified by our crystal structure are conserved in the PRORP2 and PRORP3 isoforms but some, including Tyr133, are different. Tyr133 is Gln in PRORP2 and PRORP3 and in the more distantly related human mitochondrial MRPP3, it is Asp (Figure 3-8). To investigate the importance of these interactions we measured the binding affinity of the PRORP1 Y133Q and Y133D mutations for 5´-fluorescein-pre-tRNA substrate using fluorescence anisotropy analysis (Figure 3-9A, Table 3-3). These substitutions decreased binding affinity by 8-fold and 28-fold relative to wild-type protein, equivalent to a loss of 1.2 and 1.9 kcal/mol, and these losses are similar to the decreased binding affinities of the Y133A (17-fold) and Y133F (30-fold) mutants that were reported previously (36).

A



B

*At*PRORP1 Residue	Protein	Residue	RNA Recognition Function
■ K101	*At*PRORP2	L34	base-stacking interaction
	*At*PRORP3	L34	
	*Hs*MRPP3	I145	
● K109	*At*PRORP2	K42	base-specific interaction
	*At*PRORP3	K42	
	*Hs*MRPP3	S153	
● Y133	*At*PRORP2	Q67	base-specific interaction
	*At*PRORP3	Q67	
	*Hs*MRPP3	D176	
■ Y140	*At*PRORP2	Y74	base-stacking interaction
	*At*PRORP3	Y74	
	*Hs*MRPP3	Y183	
▼ R184	*At*PRORP2	R117	phosphate-salt bridge interaction
	*At*PRORP3	R117	
	*Hs*MRPP3	R218	
● R210	*At*PRORP2	R145	base-specific interaction
	*At*PRORP3	R145	
	*Hs*MRPP3	S243	
▼ R212	*At*PRORP2	R147	phosphate-salt bridge interaction
	*At*PRORP3	R147	
	*Hs*MRPP3	K245	

Figure 3-7 Conservation of tRNA-interacting residues across the PRORP family.

(A) Amino acid sequence alignment of the PPR domain from members of the PRORP family in *A. thaliana* and human MRPP3. Higher conservation is indicated by a darker background. Red closed-circles indicate residues involved in base-specific interactions in PRORP1. Blue boxes and triangles indicate residues forming base-stacking and phosphate-salt bridge interactions,

respectively. (**B**) tRNA structure specific residues are well conserved across the PRORP family. Amino acid residues in *A. thaliana* PRORP2 and PRORP3 and human MRPP3 at equivalent positions to the tRNA-interacting residues of *A. thaliana* PRORP1. The amino acid residue numbering for PRORP3 matches that in Gobert, et al. (14) and Brillante, et al. (37).

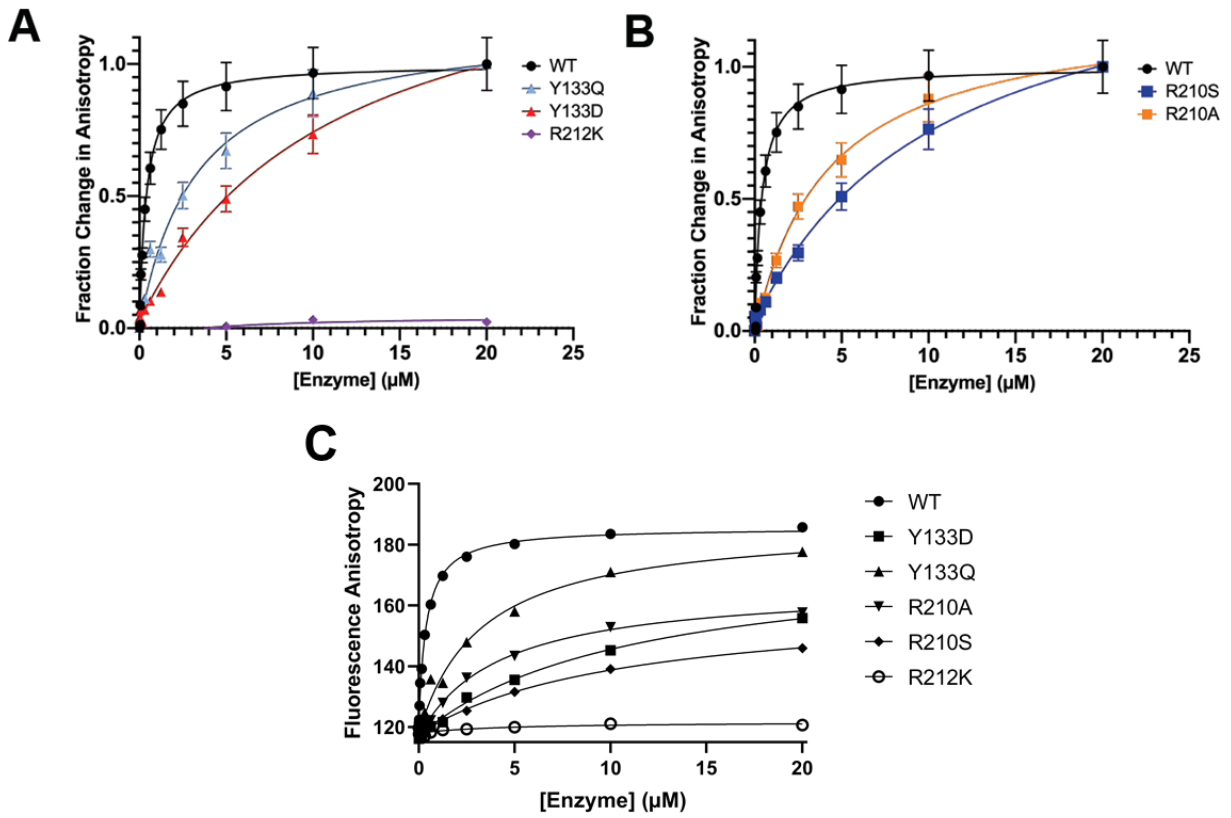


Figure 3-8 Fluorescence anisotropy binding curves for PRORP1 variants binding to *B. subtilis* 5´-fluorescein-pre-tRNA^{Asp} substrate.

(A) Binding curves for PRORP1 Y133 and R212 variants. **(B)** Binding curves for PRORP1 R210 variants. WT binding curves are shown in black. The assays were carried out in 30 mM MOPS, pH 7.8, 330 mM NaCl, 1 mM TCEP, and 20 mM $CaCl_2$ at 25 ± 1 °C. A hyperbola (Eq. 1, Materials and Methods) was fit to the fraction change in anisotropy measured from a single experimental trial using GraphPad Prism to derive values for the dissociation constant (K_D) (Table 3-3). **(C)** Fitting of the non-normalized raw anisotropy data, showing differences in total change in anisotropy. Both normalize and non-normalized fits gave the same K_D values.

In addition to Tyr133, hydrogen bond interactions between Arg210 and Lys109 and the tRNA appeared to be base specific. Arg210 interacts with the O2 group of ψ55 in our crystal structure (Figures 3-4B and 6) and it is conserved in PRORPs 1- (Figure 3-8). To evaluate its

involvement in tRNA recognition, we generated two mutants, R210A and R210S, and measured their pre-tRNA binding affinity by fluorescence anisotropy analysis. We found that the R210A and R210S mutations increased the pre-tRNA K_D values by 11-fold and 27-fold, respectively, compared to that of wild-type PRORP1 (Figure 3-9B, Table 3-3). These losses in binding affinity correspond to decreases of 1.4 and 1.9 kcal/mol, respectively. This result demonstrates that the base-specific interaction of Arg210 in the PRORP1 PPR3 plays a crucial role in tRNA recognition. Unfortunately, we could not directly test the effect of a base substitution in the tRNA because ψ55 also interacts with G18 and the phosphate backbone such that base substitutions would also affect the tRNA structure. Lys109 contacts the O4 atom of D17, a modified nucleotide (Figure 3-4B), however mutation to alanine was shown previously to have little effect on RNA-binding affinity (Table 3-4) (40). Lys109 is conserved among PRORPs 1-3 (Figure 3-8), but the nucleotide at position 17 in the tRNA D loop is variable (Figure 3-10). The modest effect of the K109A mutation appears consistent with the need to tolerate tRNA substitutions.

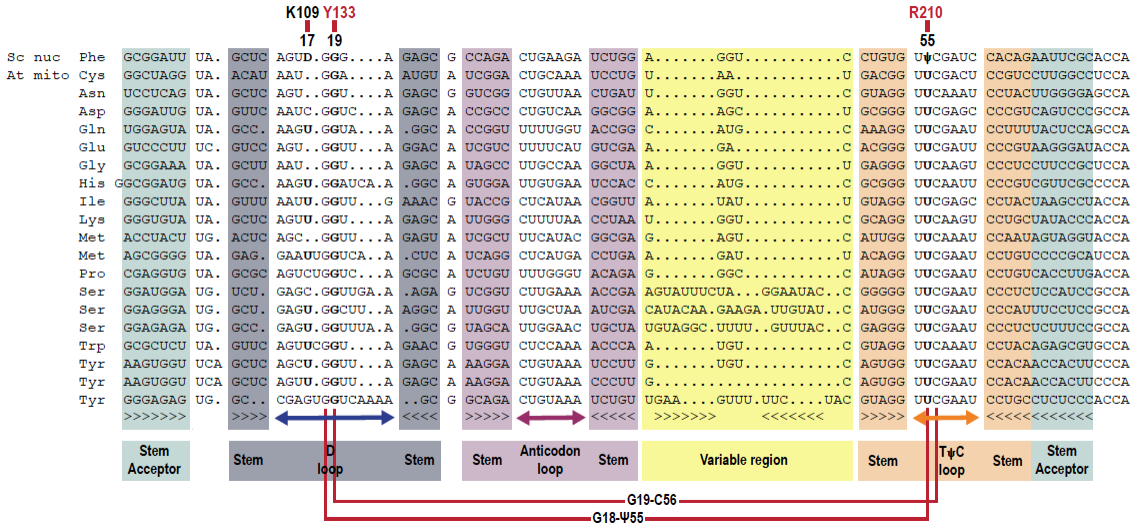


Figure 3-9 Nucleotide sequence alignment of yeast tRNA^{Phe} and *Arabidopsis* mitochondrial tRNAs.

To further understand the role that the newly identified tRNA-interacting residues (Tyr133, Arg210, and Arg212) play in catalysis, single-turnover kinetic rate constants (k_{obs}) were determined under enzyme saturating conditions with limiting (30 nM) 5´-fluorescein-pre-tRNA^{Asp} substrate (Table 3-3 and Figure 3-11). Under saturating conditions, k_{obs} measures the reaction of the enzyme-bound substrate, removing the effects of altering the binding affinity of the substrate. Therefore, significant changes in k_{obs} compared to wild-type PRORP1 indicate that the mutated residue plays a role in more than just the initial binding step. Of the isoforms measured, Y133Q maintained wild-type level k_{obs} (Table 3-3), indicating that the Y133Q mutation interfered with equilibrium association but not catalytic competence. R212K displayed appreciable loss in binding affinity ($K_D \geq 50$ μM), suggesting that this residue is important for tRNA recognition. However, because we were unable to achieve enzyme saturating conditions for R212K, we cannot accurately determine the single-turnover activity for this mutant enzyme to compare with wild-type (Figure 3-11).

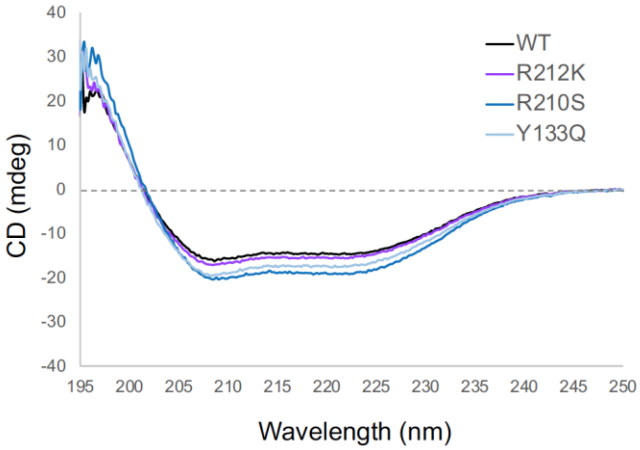


Figure 3-10 CD spectra of full-length PRORP1 wild-type and mutant proteins.

Spectra were measured at 20 ÅãC with protein concentration of 0.13 mg/ml in 25 mM Na phosphate, pH 7.5, 200 mM NaCl. The raw CD data were adjusted by subtracting a buffer blank. CD spectra of wild-type and mutant proteins displayed negative ellipticities at 208/222 nm and 215 nm, which indicate the presence of α helices and β strands, respectively.

Comparison of the R212K CD spectra with wild-type and other variants confirmed that R212K secondary structural folding characteristics are retained, indicating that loss in binding affinity is due to a loss of important binding interactions.

In contrast, Y133D and R210A exhibited ~2-fold decreases in single-turnover activity while R210S displayed a ~5-fold decrease in single-turnover activity (Table 3-3 and Figure 3-11). These decreases indicate that cleavage of substrate in the enzyme-substrate complex was compromised Additionally, in our fluorescence binding assays, total change in fluorescence anisotropy for these three variants was appreciably lower than those showing wild-type single-turnover activity (Figure 3-9C). Within these assays, the positioning of the substrate fluorophore on the end of the 5' leader would result in greater total changes in anisotropy when sterically hindered by catalytically competent positioning of the metallonuclease domain. Based on these data, we suggest that the specific interactions between the pre-tRNA 'elbow' and the PRORP PPR domain mediated by Tyr133 and Arg210 play a role in optimally positioning the 5´ leader cleavage site at the metallonuclease domain active site. Mutation of these residues interferes with the formation of a catalytically competent PRORP-pre-tRNA complex as well as decreasing substrate binding affinity. Similarly, the absence of the substrate D-loop, which is recognized by Tyr133, decreases the single-turnover activity and affinity of PRORP3 (37).

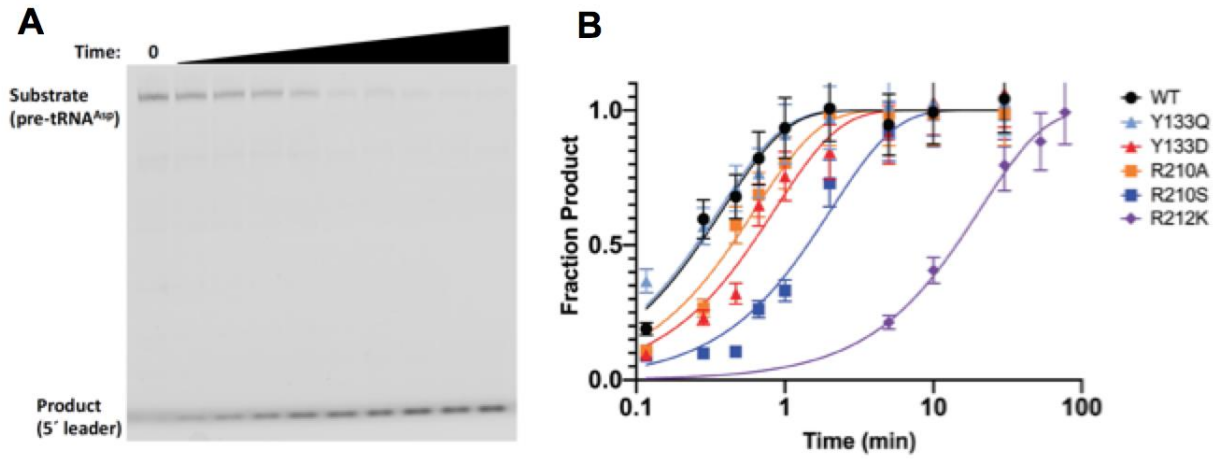


Figure 3-11 Time courses for the single-turnover cleavage of 5´-fluorescein pre-tRNA substrate catalyzed by PRORP1 variants.

(**A**) Representative single-turnover cleavage assay. (**B**) The enzyme concentration was 30 μM. A single exponential (Equation 2, Materials and Methods) was fit to the time dependence of the fraction of product formation measured from a single experimental trial using GraphPad Prism to determine a single-turnover rate constant (k_{obs}) (Table 3-3). Error shown is taken from error in gel analysis software, not the fit. The R212K mutant was evaluated under sub-saturating enzyme concentration. Data was collected in 30 mM MOPS, pH 7.8, 330 mM NaCl, 1 mM TCEP, and 20 mM $MgCl_2$ at 25 ± 1 °C.

3.5 Discussion

Our crystal structure of the PPR domain of PRORP1 in complex with tRNA illustrates its substrate recognition strategy emphasizing interactions with conserved nucleotide and structural elements of the tRNA elbow region and employing new mechanisms of PPR motif-RNA interaction. This specific recognition of the pre-tRNA elbow is essential for catalytic activity. By bringing together tRNA nucleotide and structural recognition, it requires the expansion of RNA recognition modes by PPR motifs beyond the established RNA base recognition mechanism.

The new PPR motif recognition modes in PRORP1 confirm one of the distinct proposals from the models of PPR motif-tRNA interaction proposed previously (24,36,37): that the PPR domain in PRORP does not use the canonical base-selection model, but rather utilizes a strategy like recognition of the G19-C56 base pair by the RNP RNase P ribozyme (Figure 3-12) (9).

Evolution of the RNA recognition modes of PPR motifs to allow recognition of structured RNA in addition to single-stranded RNA sequences is reminiscent of the PUF (Pumilio/*fem-3* binding factor) family of proteins. Both types of α-helical repeat proteins are well known for their modular recognition of RNA sequences (51). In addition, these α-helical repeat scaffolds can be modified for structured RNA recognition (52,53). Many other families of α-helical repeat proteins are known for their roles in RNA metabolism (54), presenting the possibility that they, too, serve as versatile scaffolds for RNA recognition.

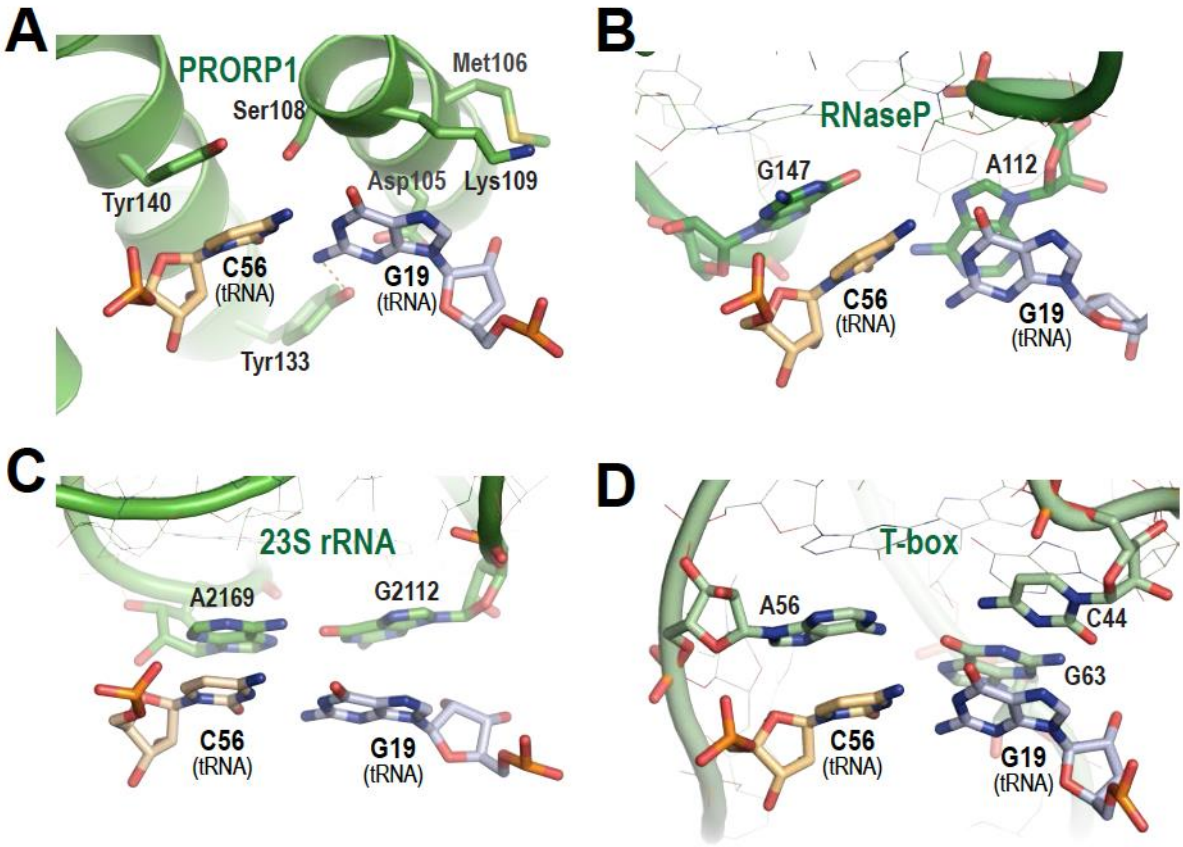


Figure 3-12 Evolutionary convergence of G19-C56 tRNA base pair recognition.

(**A**) PRORP1 PPR domain recognition of the G19-C56 base pair. (**B**) Bacterial RNP RNase P RNA recognition of the G19-C56 base pair (PDB ID: 3Q1Q). (**C**) 23S rRNA recognition of the G19-C56 base pair (PDB ID: 4V4I). (**D**) T-box riboswitch recognition of the G19-C56 base pair (PDB ID: 4LCK). PRORP1 PPR domain, RNase P RNA, 23S rRNA, and T-box riboswitch in green are shown as cartoons with interacting residues displayed as stick models. Amino acid side chains and tRNA D-loop (light blue) and TψC-loop (orange) nucleotides are shown with atom colors: oxygen

(red), nitrogen (blue), phosphorus (orange) and sulfur (yellow). Non-tRNA interacting nucleotides of bacterial RNase P, 23S rRNA, and T-box riboswitch RNAs are shown as thin stick models.

The central features of recognition of the tRNA elbow are a stacking interaction between PRORP1 Tyr140 and C56 in the G19-C56 base pair and electrostatic interactions of Arg184 and Arg212 with phosphate groups important for the positioning of G57. These highly conserved tRNA structural elements are recognized by residues that are retained in *Arabidopsis* PRORPs and the human MRPP3 (Figure 3-8). Mutation of these residues has severe effects on pre-tRNA binding affinity (Table 3-4). Mutation of Tyr140 to alanine in full-length PRORP1 severely diminished pre-tRNA binding affinity (~200-fold weaker) and mutation to phenylalanine decreased pre-tRNA binding affinity ~6-fold (36). Similarly, mutation of Arg184 and Arg212 to alanine dramatically reduced pre-tRNA binding affinity. In addition, base-specific interactions of PRORP1 Tyr133 with the N2 atom of G19 and Arg210 with the O4 atom of ψ55 appear to be important for pre-tRNA recognition. Mutation of Tyr133 or Arg210 decreases pre-tRNA binding affinity, indicating their important contribution to recognition. Despite their importance in PRORP1, Tyr133 is not conserved in *Arabidopsis* PRORP2 and PRORP3, and neither Tyr133 nor Arg210 is conserved in human MRPP3 (Figure 3-8). PRORP2 and PRORP3, which process pre-tRNAs retaining the G19-C56 base pair between the D and TψC loops, may have evolved alternative mechanisms for base-specific binding consistent with their set of pre-tRNA substrates. For MRPP3, mammalian mitochondrial tRNAs retain the overall L-shape, but their sequences are degenerate, especially in the D and TψC loops. Consequently, MRPP3 may recognize only the structure of pre-tRNA substrates.

Our crystal structure indicates conformational changes in PRORP enzymes that are required for recognition and 5´ cleavage of pre-tRNA. Previous biochemical and biophysical

analyses suggested that PRORPs adopt multiple conformations (35). A hinge between the NYN metallonuclease and central zinc-binding domains allows reorientation of the nuclease domain with respect to the PPR domain. Superposition of our tRNA-bound PPR domain with the PPR domains in crystal structures of *Arabidopsis* PRORP1 or PRORP2 shows that without a conformational change the position of the nuclease domain would overlap the acceptor-stem of the tRNA (Figure 3-13). Therefore, the relative positions of the nuclease and PPR domains must be more open to recognize the pre-tRNA elbow and place the nuclease active site at the pre-tRNA 5´ end. Molecular dynamics simulations have illustrated the ranges of conformational flexibility in PRORP enzymes (35,55). In addition to this requirement for overall conformational change, we observed flexibility within the PRORP PPR domain (Figure 3-3) that may be crucial to recognize pre-tRNA and to release mature tRNA product.

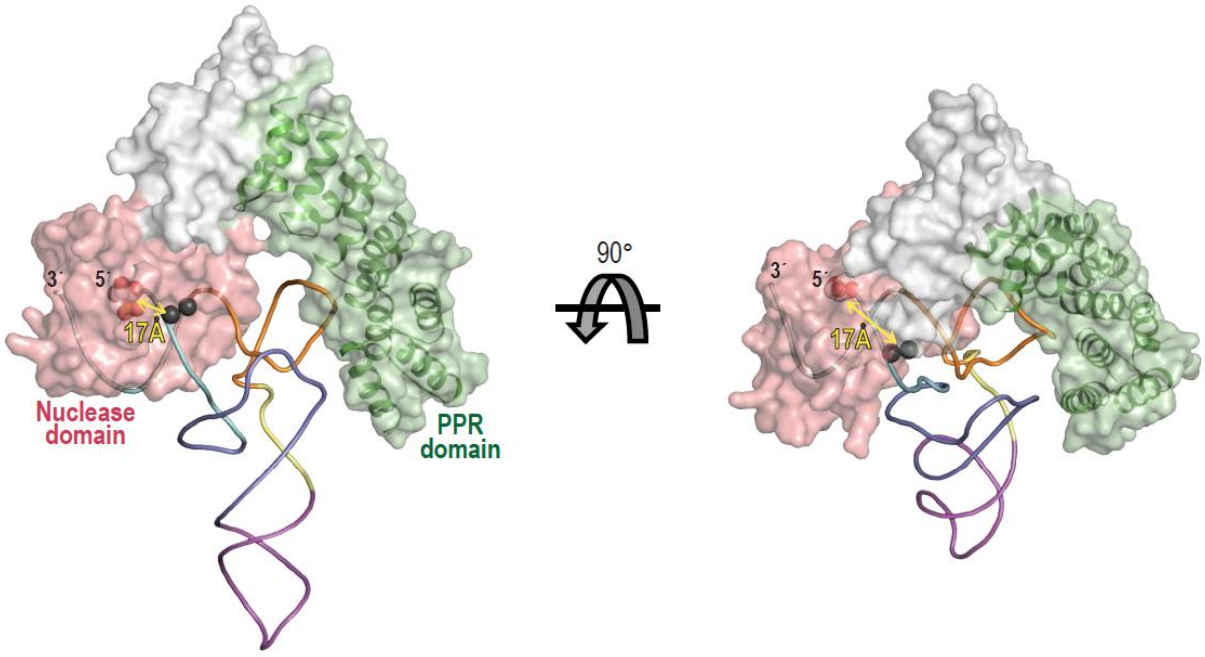


Figure 3-13 Superposition of crystal structures of the PRORP1 PPR-tRNA complex and fulllength PRORP1.

The PPR domain and tRNA of the complex are shown as ribbon diagram and cartoon representation, respectively. Full-length PRORP1 (PDB ID: 4G24) is shown as a transparent

surface with green PPR domain, gray central zinc binding domain, and pink NYN metallonuclease domain. The tRNA is shown as a backbone trace colored by region: acceptor stem (cyan), D stem loop (blue), anticodon stem loop (magenta), variable region (yellow), and TψC stem loop (orange). The distance between the 5ÅL end of the tRNA backbone indicating the PRORP cleavage site (shown with space-filling red spheres) and the catalytic metal ions (black spheres) marking the nuclease active site is indicated in yellow.

PRORP enzymes must process a set of diverse pre-tRNA substrates, and our data suggest multiple mechanisms to promote recognition plasticity. In addition to tRNA sequence diversity, 5′ leader and 3′ trailer processing occur prior to structure-stabilizing base modifications (56,57). Therefore, PRORP enzymes might recognize and bind partially folded or dynamic pre-tRNA tertiary structures during the early stages of maturation, and in turn they might aid in achieving the final tRNA tertiary structure. Recognition of a highly conserved pre-tRNA structural feature like the G19-C56 base pair is an excellent strategy to capture a range of pre-tRNA conformations. Our data indicating that the PRORP1 PPR domain binds equally well to modified and unmodified tRNA also allows binding as the tRNA matures. Flexibility of the PPR domain and overall PRORP1 conformation permits optimization of binding to different pre-tRNA structures.

Recognition of the G19-C56 base pair at the tRNA elbow is a core element of PRORP PPR interactions, and it is striking that this mode of recognition by a protein is analogous to recognition of the same feature by three functional RNAs (Figure 3-12). The crystal structure of *Thermotoga maritima* RNP RNase P in complex with tRNA showed that highly conserved nucleotides A112 and G147 in single-stranded loops of the RNase P RNA form a binding pocket and interact via stacking interactions with G19 and C56 (Figure 3-12B) (9). This interaction is conserved in structures of RNP RNase P from bacteria, archaea, yeast, and human (Figure 3-14) (8-11). Furthermore, this tRNA-binding mode is like that between the L1 stalk of the 23S rRNA and tRNA at the E-site in the 50S ribosome (38) and two single-stranded loops of T-box riboswitches and their cognate tRNAs (39) (Figure 3-12C, D). Since these functional RNAs share no common

ancestor, it is thus postulated that the specific binding to the tRNA elbow with stacking interaction has evolved independently at least three times (39). Evolution of the recognition α-helices in the PPR1 and PPR2 motifs mimics the binding pocket for the tRNA elbow formed by the two single-stranded chains of these functional RNAs. Hence, PRORPs add a fourth molecule to this evolutionarily convergent solution for tRNA elbow recognition.

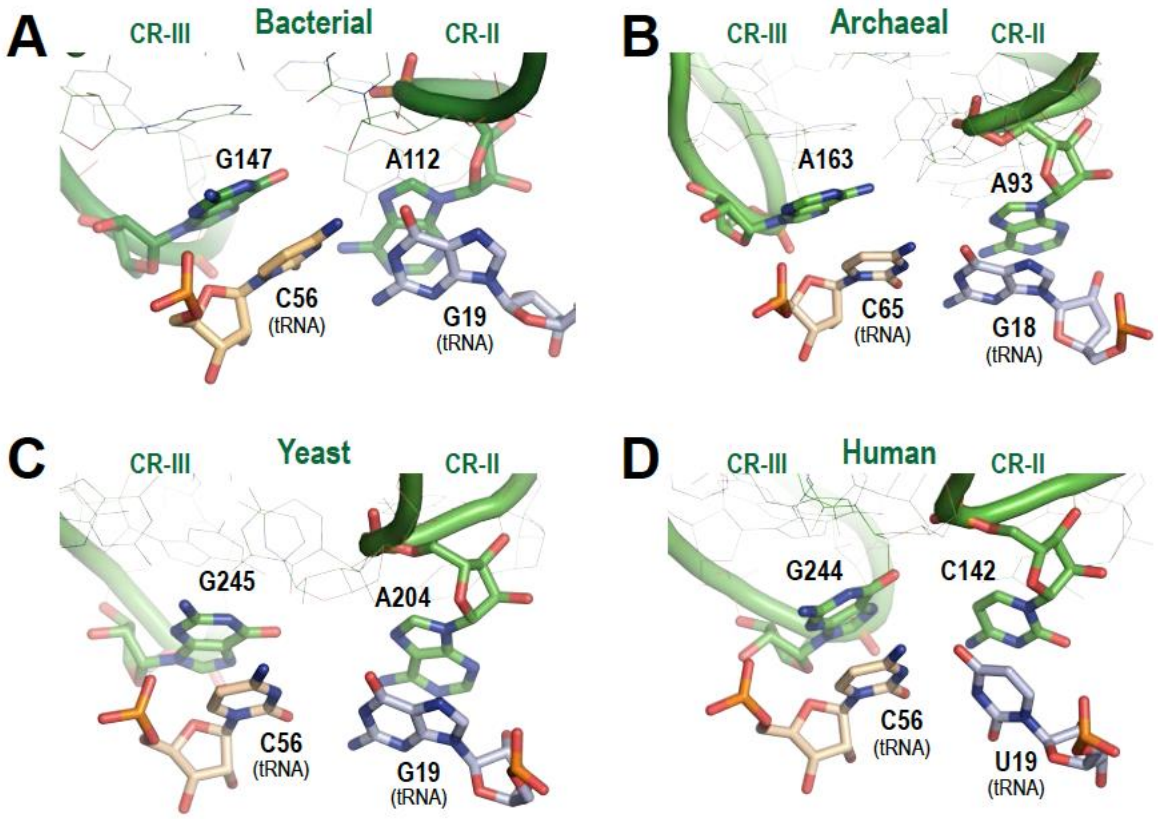


Figure 3-14 Recognition of the tRNA elbow base pair by nucleotides in the CR-II and CR-III regions of the RNase P RNA is conserved across kingdoms.

(**A**) Bacterial RNP RNase P (PDB ID:3Q1Q). (**B**) Archaeal RNP RNase P (PDB ID: 6K0B). (**C**) Yeast RNP RNase P (PDB ID: 6AH3). (**D**) Human RNP RNase P (PDB ID: 6AHU). RNase P RNA is shown in green as cartoons with interacting nucleotides displayed as stick models. tRNA D-loop (light blue) and TψC-loop (orange) nucleotides are shown with atom colors: oxygen (red), nitrogen (blue), and phosphorus (orange). Non-tRNA interacting nucleotides of RNase P RNAs are shown as thin stick models.

3.6 References

1. Abbott, J.A., Francklyn, C.S. and Robey-Bond, S.M. (2014) Transfer RNA and human disease. *Front Genet*, **5**, 158.
2. Esakova, O. and Krasilnikov, A.S. (2010) Of proteins and RNA: the RNase P/MRP family. *RNA*, **16**, 1725-1747.
3. Schelcher, C., Sauter, C. and Giege, P. (2016) Mechanistic and Structural Studies of Protein-Only RNase P Compared to Ribonucleoproteins Reveal the Two Faces of the Same Enzymatic Activity. *Biomolecules*, **6**.
4. Daniels, C.J., Lai, L.B., Chen, T.H. and Gopalan, V. (2019) Both kinds of RNase P in all domains of life: surprises galore. *RNA*, **25**, 286-291.
5. Nickel, A.I., Waber, N.B., Gossringer, M., Lechner, M., Linne, U., Toth, U., Rossmanith, W. and Hartmann, R.K. (2017) Minimal and RNA-free RNase P in Aquifex aeolicus. *Proc Natl Acad Sci U S A*, **114**, 11121-11126.
6. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, **35**, 849-857.
7. Kazantsev, A.V. and Pace, N.R. (2006) Bacterial RNase P: a new view of an ancient enzyme. *Nat Rev Microbiol*, **4**, 729-740.
8. Lan, P., Tan, M., Zhang, Y., Niu, S., Chen, J., Shi, S., Qiu, S., Wang, X., Peng, X., Cai, G. *et al.* (2018) Structural insight into precursor tRNA processing by yeast ribonuclease P. *Science*, **362**.
9. Reiter, N.J., Osterman, A., Torres-Larios, A., Swinger, K.K., Pan, T. and Mondragon, A. (2010) Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature*, **468**, 784-789.
10. Wan, F., Wang, Q., Tan, J., Tan, M., Chen, J., Shi, S., Lan, P., Wu, J. and Lei, M. (2019) Cryo-electron microscopy structure of an archaeal ribonuclease P holoenzyme. *Nat Commun*, **10**, 2617.
11. Wu, J., Niu, S., Tan, M., Huang, C., Li, M., Song, Y., Wang, Q., Chen, J., Shi, S., Lan, P. *et al.* (2018) Cryo-EM Structure of the Human Ribonuclease P Holoenzyme. *Cell*, **175**, 1393-1404 e1311.
12. Holzmann, J., Frank, P., Loffler, E., Bennett, K.L., Gerner, C. and Rossmanith, W. (2008) RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell*, **135**, 462-474.
13. Vilardo, E., Nachbagauer, C., Buzet, A., Taschner, A., Holzmann, J. and Rossmanith, W. (2012) A subcomplex of human mitochondrial RNase P is a bifunctional methyltransferase--extensive moonlighting in mitochondrial tRNA biogenesis. *Nucleic Acids Res*, **40**, 11583-11593.
14. Gobert, A., Gutmann, B., Taschner, A., Gossringer, M., Holzmann, J., Hartmann, R.K., Rossmanith, W. and Giege, P. (2010) A single Arabidopsis organellar protein has RNase P activity. *Nat Struct Mol Biol*, **17**, 740-744.
15. Gutmann, B., Gobert, A. and Giege, P. (2012) PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis. *Genes Dev*, **26**, 1022-1027.
16. Lai, L.B., Bernal-Bayard, P., Mohannath, G., Lai, S.M., Gopalan, V. and Vioque, A. (2011) A functional RNase P protein subunit of bacterial origin in some eukaryotes. *Mol Genet Genomics*, **286**, 359-369.

17. Taschner, A., Weber, C., Buzet, A., Hartmann, R.K., Hartig, A. and Rossmanith, W. (2012) Nuclear RNase P of Trypanosoma brucei: a single protein in place of the multicomponent RNA-protein complex. *Cell Rep*, **2**, 19-25.
18. Sugita, C., Komura, Y., Tanaka, K., Kometani, K., Satoh, H. and Sugita, M. (2014) Molecular characterization of three PRORP proteins in the moss Physcomitrella patens: nuclear PRORP protein is not essential for moss viability. *PLoS One*, **9**, e108962.
19. Lechner, M., Rossmanith, W., Hartmann, R.K., Tholken, C., Gutmann, B., Giege, P. and Gobert, A. (2015) Distribution of Ribonucleoprotein and Protein-Only RNase P in Eukarya. *Mol Biol Evol*, **32**, 3186-3193.
20. Bouchoucha, A., Waltz, F., Bonnard, G., Arrive, M., Hammann, P., Kuhn, L., Schelcher, C., Zuber, H., Gobert, A. and Giege, P. (2019) Determination of protein-only RNase P interactome in Arabidopsis mitochondria and chloroplasts identifies a complex between PRORP1 and another NYN domain nuclease. *Plant J*, **100**, 549-561.
21. Weber, C., Hartig, A., Hartmann, R.K. and Rossmanith, W. (2014) Playing RNase P evolution: swapping the RNA catalyst for a protein reveals functional uniformity of highly divergent enzyme forms. *PLoS Genet*, **10**, e1004506.
22. Howard, M.J., Lim, W.H., Fierke, C.A. and Koutmos, M. (2012) Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5' processing. *Proc Natl Acad Sci U S A*, **109**, 16149-16154.
23. Karasik, A., Shanmuganathan, A., Howard, M.J., Fierke, C.A. and Koutmos, M. (2016) Nuclear Protein-Only Ribonuclease P2 Structure and Biochemical Characterization Provide Insight into the Conserved Properties of tRNA 5' End Processing Enzymes. *J Mol Biol*, **428**, 26-40.
24. Gobert, A., Pinker, F., Fuchsbauer, O., Gutmann, B., Boutin, R., Roblin, P., Sauter, C. and Giege, P. (2013) Structural insights into protein-only RNase P complexed with tRNA. *Nat Commun*, **4**, 1353.
25. Fujii, S. and Small, I. (2011) The evolution of RNA editing and pentatricopeptide repeat genes. *New Phytol*, **191**, 37-47.
26. Kotera, E., Tasaka, M. and Shikanai, T. (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature*, **433**, 326-330.
27. Nakamura, T., Yagi, Y. and Kobayashi, K. (2012) Mechanistic insight into pentatricopeptide repeat proteins as sequence-specific RNA-binding proteins for organellar RNAs in plants. *Plant Cell Physiol*, **53**, 1171-1179.
28. Ringel, R., Sologub, M., Morozov, Y.I., Litonin, D., Cramer, P. and Temiakov, D. (2011) Structure of human mitochondrial RNA polymerase. *Nature*, **478**, 269-273.
29. Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y.S., Bond, C.S. and Small, I. (2012) A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet*, **8**, e1002910.
30. Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T. and Nakamura, T. (2013) Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS One*, **8**, e57286.
31. Yagi, Y., Tachikawa, M., Noguchi, H., Satoh, S., Obokata, J. and Nakamura, T. (2013) Pentatricopeptide repeat proteins involved in plant organellar RNA editing. *RNA Biol*, **10**, 1419-1425.

32. Shen, C., Zhang, D., Guan, Z., Liu, Y., Yang, Z., Yang, Y., Wang, X., Wang, Q., Zhang, Q., Fan, S. *et al.* (2016) Structural basis for specific single-stranded RNA recognition by designer pentatricopeptide repeat proteins. *Nat Commun*, **7**, 11285.
33. Yin, P., Li, Q., Yan, C., Liu, Y., Liu, J., Yu, F., Wang, Z., Long, J., He, J., Wang, H.W. *et al.* (2013) Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature*, **504**, 168-171.
34. Imai, T., Nakamura, T., Maeda, T., Nakayama, K., Gao, X., Nakashima, T., Kakuta, Y. and Kimura, M. (2014) Pentatricopeptide repeat motifs in the processing enzyme PRORP1 in Arabidopsis thaliana play a crucial role in recognition of nucleotide bases at TpsiC loop in precursor tRNAs. *Biochem Biophys Res Commun*, **450**, 1541-1546.
35. Pinker, F., Schelcher, C., Fernandez-Millan, P., Gobert, A., Birck, C., Thureau, A., Roblin, P., Giege, P. and Sauter, C. (2017) Biophysical analysis of Arabidopsis protein-only RNase P alone and in complex with tRNA provides a refined model of tRNA binding. *J Biol Chem*, **292**, 13904-13913.
36. Klemm, B.P., Karasik, A., Kaitany, K.J., Shanmuganathan, A., Henley, M.J., Thelen, A.Z., Dewar, A.J.L., Jackson, N.D., Koutmos, M. and Fierke, C.A. (2017) Molecular recognition of pre-tRNA by Arabidopsis protein-only Ribonuclease P. *RNA*, **23**, 1860-1873.
37. Brillante, N., Gossringer, M., Lindenhofer, D., Toth, U., Rossmanith, W. and Hartmann, R.K. (2016) Substrate recognition and cleavage-site selection by a single-subunit protein-only RNase P. *Nucleic Acids Res*, **44**, 2323-2336.
38. Korostelev, A., Trakhanov, S., Laurberg, M. and Noller, H.F. (2006) Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell*, **126**, 1065-1077.
39. Zhang, J. and Ferre-D'Amare, A.R. (2013) Co-crystal structure of a T-box riboswitch stem I domain in complex with its cognate tRNA. *Nature*, **500**, 363-366.
40. Chen, T.H., Tanimoto, A., Shkriabai, N., Kvaratskhelia, M., Wysocki, V. and Gopalan, V. (2016) Use of chemical modification and mass spectrometry to identify substrate-contacting sites in proteinaceous RNase P, a tRNA processing enzyme. *Nucleic Acids Res*, **44**, 5344-5355.
41. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*, **276**, 307-326.
42. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, **60**, 2126-2132.
43. Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*, **53**, 240-255.
44. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*, **66**, 213-221.
45. Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, **66**, 12-21.
46. Howard, M.J., Klemm, B.P. and Fierke, C.A. (2015) Mechanistic Studies Reveal Similar Catalytic Strategies for Phosphodiester Bond Hydrolysis by Protein-only and RNA-dependent Ribonuclease P. *J Biol Chem*, **290**, 13454-13464.

47. Milligan, J.F. and Uhlenbeck, O.C. (1989) Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol*, **180**, 51-62.
48. Grove, T.Z., Cortajarena, A.L. and Regan, L. (2008) Ligand binding by repeat proteins: natural and designed. *Curr Opin Struct Biol*, **18**, 507-515.
49. Coquille, S., Filipovska, A., Chia, T., Rajappa, L., Lingford, J.P., Razif, M.F., Thore, S. and Rackham, O. (2014) An artificial PPR scaffold for programmable RNA recognition. *Nat Commun*, **5**, 5729.
50. Shi, H. and Moore, P.B. (2000) The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited. *RNA*, **6**, 1091-1105.
51. Hall, T.M.T. (2016) De-coding and re-coding RNA recognition by PUF and PPR repeat proteins. *Curr Opin Struct Biol*, **36**, 116-121.
52. Qiu, C., McCann, K.L., Wine, R.N., Baserga, S.J. and Hall, T.M.T. (2014) A divergent Pumilio repeat protein family for pre-rRNA processing and mRNA localization. *Proc Natl Acad Sci U S A*, **111**, 18554-18559.
53. Zhang, J., McCann, K.L., Qiu, C., Gonzalez, L.E., Baserga, S.J. and Hall, T.M.T. (2016) Nop9 is a PUF-like protein that prevents premature cleavage to correctly process pre-18S rRNA. *Nat Commun*, **7**, 13085.
54. Hammani, K., Bonnard, G., Bouchoucha, A., Gobert, A., Pinker, F., Salinas, T. and Giege, P. (2014) Helical repeats modular proteins are major players for organelle gene expression. *Biochimie*, **100**, 141-150.
55. Chen, T.H., Sotomayor, M. and Gopalan, V. (2019) Biochemical Studies Provide Insights into the Necessity for Multiple Arabidopsis thaliana Protein-Only RNase P Isoenzymes. *J Mol Biol*, **431**, 615-624.
56. Binder, S., Stoll, K. and Stoll, B. (2016) Maturation of 5' ends of plant mitochondrial RNAs. *Physiol Plant*, **157**, 280-288.
57. Rackham, O., Busch, J.D., Matic, S., Siira, S.J., Kuznetsova, I., Atanassov, I., Ermer, J.A., Shearwood, A.M., Richman, T.R., Stewart, J.B. *et al.* (2016) Hierarchical RNA Processing Is Required for Mitochondrial Ribosome Assembly. *Cell Rep*, **16**, 1874-1890.

Chapter 4 Conclusions and Future Directions

4.1 Conclusions

The goal of this work was to expand our understanding of the substrate recognition strategies of PRORP, particularly focusing on contributions of the PPR domain. Biochemical and structural investigations revealed that the mechanism of substrate recognition in PRORP is analogous to the RNase P ribozyme and identified a new mechanism of PPR motif RNA interaction different from that of sequence-specific ssRNA binding proteins.

Protein-Only RNase P Recognition

Site-directed mutagenesis and kinetic and thermodynamic experiments were used to find interactions with PRORP residues important for pre-tRNA recognition. These results identified several residues within the concave surface of the PPR domain as significant contributors to pre-tRNA recognition. We also used sodium ion dependence of substrate binding affinity demonstrating the importance of ionic interactions in pre-tRNA recognition by PRORP. Together the data from this biochemical analysis was used to produce a predictive structural model of PRORP-substrate recognition. Our X-ray crystal structure of tRNA bound to the PRORP PPR domain largely validated this recognition model, showing that PPR domain residues form specific interactions with conserved tRNA bases and structural regions. Taken together this work suggests that PRORP has evolved to utilize a similar pre-tRNA recognition strategy as the RNA-based enzyme. Additionally, we identified a new mode of RNA recognition by PPR motifs, thus expanding the PPR-RNA interaction beyond the established sequence-

specific ssRNA mode. This recognition mechanism may turn out to be used by other PPR proteins as well.

4.2 Future Directions

The work described in these chapters has greatly expanded our understanding of PRORP substrate recognition, nonetheless many questions remain unsolved. Firstly, it is unclear interactions with the PPR contribute to enhancement of the rate of the cleavage step remains unclear. Previous studies, as well as ours, have indicated that conformational changes are involved in PRORP catalysis, yet a mechanistic basis for that predicted contribution is unknown. Second, is the substrate recognition strategy observed in PRORP1 shared amongst all the PRORP homologs? Though metazoan mitochondrial PRORP (i.e. human MRPP3) is analogous in structure, two additional protein subunits are required to activate catalysis and the role of these proteins in substrate recognition is unknown. Additionally, many pre-tRNA substrates of these PRORP homologues lack the conserved nucleotides which are recognized by single-subunit PRORP. Lastly, the recent discovery of a largely uncharacterized novel from of PRORP within bacteria and Archaea, offer the opportunity to further expand our understanding of the function and evolution of RNase P.

4.2.1 PRORP Conformational Changes

Our kinetic and thermodynamic data indicate that interactions with the PPR motifs in PRORP enhance thermodynamic substrate association and the cleavage rate constant, possibly by stabilizing alternate conformations of the PRORP-pre-tRNA complex. Additionally, conformational changes in PRORP are believed to be required for recognition and 5’ leader cleavage of pre-tRNA. Our crystal structure indicates that there are changes in the PPR motif structure upon binding, while our model suggests that the overall V shape topology must open to

fit the bound substrate. This may indicate that like RNA-based RNase P, an induced-fit mechanism of recognition is present in PRORP catalysis, initial binding of PRORP to the tRNA body induces a conformational change in PRORP to form a catalytically competent PRORP-pre-tRNA complex. Transient kinetic stopped-flow experiments and the pursuit of a structure of full-length PRORP bound to substrate are approaches that could be used to further probe the recognition mechanism of PRORP. Stopped-flow kinetic experiments with PRORP and 5' fluorescently labeled pre-tRNA substrate could be used to detect multiple exponentials in the formation of the PRORP-pre-tRNA complex indicative of conformational change step that occurs between biomolecular association and cleavage. Initial stopped-flow experiments that I carried out showed multiple phases upon mixing PRORP with fluorescent-labeled pre-tRNA, consistent with one or more conformational transitions. Additional detailed studies of the kinetics of binding would need to be completed to identify steps linked with association, conformational change and substrate cleavage. If a conformational change were identified through these kinetic studies, a combination of site-directed mutagenesis with stopped-flow kinetics could relate the conformational change phase to specific residue interactions and thereby help elucidate the mechanistic contributions of PPR motif interactions in endonucleolytic cleavage.

Additionally, a structure of full-length PRORP bound to tRNA would help to further uncover the role of the PPR domains conformational change upon substrate binding. A comparison of our PPR-tRNA structure to the full-length unbound structure shows the concave binding surface of the PPR domain forming a more acute angle which conforms to the tRNA elbow region. Additionally, an overlay of the full-length PRORP structure, through alignment of the PPR domain, positions the metallonuclease domain 17 Ångstroms away from the scissile

phosphodiester bond (Figure 3-13). Given this, it seems likely that the tRNA induced conformational change in the PPR domain could cause PRORP structural changes that position the active site for cleavage. Elucidation of a full-length PRORP structure bound to tRNA/pre-tRNA could reveal whether this is the case, and if so, how the substrate binding induced change in PPR conformation causes an overall change in PRORP global architecture.

4.2.2 Multi-subunit PRORP Substrate Recognition

Due to the presence of two additional protein subunits and noncanonical pre-tRNA substrates, it is believed that Metazoan mitochondrial PRORP substrate recognition likely differs from single-subunit PRORP. Recently, pre-published kinetic experiments by my colleagues on human mitochondrial RNase P (mtRNase P) support this divergence in substrate recognition strategy (1). Biochemical investigation of the substrate recognition role of the stable subcomplex, MRPP1/2 (TRMT10C/SDR5C1), led Xin Liu and others to propose a minimal kinetic mechanism of mtRNase P (MRPP1/2/3) (Figure 4-1). An important discovery that led to the development of this scheme was that only in the presence of pre-tRNA is a stable MRPP1/2/3 complex observed. Within this putative kinetic mechanism, there are three distinct orders of protein subunit and pre-tRNA binding which lead to a catalytically competent conformer. Although dependent on relative concentrations of mitochondrial RNase P subcomplex components, one pathway appears to be kinetically favored. In the scenario that pre-tRNA substrate is in excess of the mitochondrial RNase P components (MRPP1/2 and MRPP3), the thermodynamic binding equilibriums indicate that the product formation would primarily take place through initial complexation of MRPP1/2*S. This kinetic model is supported by recent studies on human mitochondrial transcript processing that have indicated that the MRPP1/2 subcomplex acts as a platform for recognition by not only RNase P but additional tRNA

maturation processing enzymes. This model predicts that mitochondrial tRNA maturation takes place starting with pre-tRNA saturation with MRPP1/2 subcomplex in order to undergo recognition and catalysis by RNase P as well as other downstream tRNA processing enzymes (2). The MRPP1/MRPP2 complex is a tRNA-maturation platform in human mitochondria. However, whether the recognition mechanism of MRPP1/2-S* is due to protein-protein interaction or pre-tRNA structural rearrangement is unknown. Currently, my colleague Vojc Kocman, of Markos Koutmos lab, is working to clarify the mechanism behind this enhanced recognition using a variety of biophysical studies.

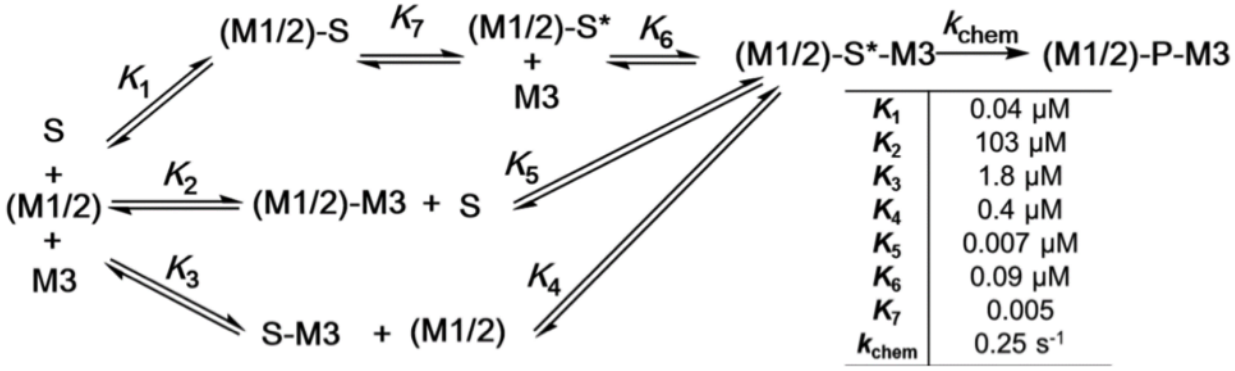


Figure 4-1 Putative kinetic mechanism of human mitochondrial RNase P with thermodynamic equilibrium constants (K).

The predominant binding order proposes a conformational change (K_7) that enhances MRPP3 binding.

Additionally, a solved structure of the metazoan mitochondrial RNase P (MRPP1/2/3) complex with bound tRNA would offer a great deal of insight into the substrate recognition mechanism of this class of PRORP. Although this could be attempted through X-ray crystallography, difficulty with obtaining crystal structures of complete single-subunit PRORP-tRNA complex indicate this may not be a fruitful approach. Cryo-EM studies, which have shown increasing success in the solution of large complex structures, may prove to be a more productive approach. If a complete structure of this complex were attained, it would answer many of the

remaining questions surrounding the substrate recognition mechanisms of metazoan mitochondrial PRORP. Primarily, how do MRPP1/2 take part within this substrate recognition process? Although MRPP3/*Hu*PRORP contains the same overall structural architecture of the single-subunit variants, it is still unclear why these additional protein subunits are necessary for catalytic activity. Specifically, does the PPR domain of MRPP3 recognize the tRNA body of pre-tRNA in a similar mechanism to the single-subunit PRORP, or does it rely on the presence of MRPP1/2? If so, does MRPP3 form protein-protein interactions within this tertiary complex or do MRPP1/2 induce RNA structural conformations that are required for MRPP3 recognition? Answers to these questions would not only further our understanding of metazoan mitochondrial transcript processing but could potentially unveil a novel protein-RNA recognition function for PPR motifs.

4.2.3 Homologs of *Aquifex aeolicus* RNase P (HARP)

The recent discovery of a novel class of prokaryotic PRORP (HARP) provides the opportunity to further our understanding of RNase P function. In its initial characterization, evidence of a HARP metallonuclease domain which preforms catalysis utilizing a similar kinetic mechanism as eukaryotic PRORP was reported (3). Kinetic and thermodynamic metal-ion and pH experiments could further elucidate this mechanism and examine whether the PRORP two-metal ion mechanism is conserved. A kinetic study relating single turnover kinetic rates to pH could reveal a relationship of dependency that would provide evidence for a metal-hydroxide nucleophile. Furthermore, kinetic studies on divalent metal-ion dependence of steady-state kinetic parameters could seek to identify cooperative divalent metal-ion binding indicating the presence of metal-ions coordinated in the active-site.

Little is known about HARP pre-tRNA recognition, as these proteins lack an identifiable RNA binding domain. Therefore, applying X-ray crystallography to solve a structure of HARP could elucidate a novel mechanism of substrate binding, presumably RNA. Even if a solved structure does not contain bound substrate, structural information of HARP alone would inform site-directed mutagenesis studies that could provide information about potential RNA binding domains.

Finally, although HARP exhibits pre-tRNA processing activity *in vitro,* it has been demonstrated that they are not the major source of RNase P activity in two Archaea (4). This leaves their functional role within these hosts in question. Therefore, their native substrates could be sought through high-throughput sequence analysis of organellar RNA, comparing HARP knockout and wild-type cell lines.

4.3 Implications of Work on the Origins of Life

Similarities in catalytic strategies and rates of RNA-dependent and protein-only RNase P seem to suggest that the transition from RNA to protein catalysts was not due mainly due to catalytic efficiency. One possible reason for the transition could be due to a need for greater stability of the catalyst. This is supported by the presence of PRORP in organisms/organelles that have harsher environments (oxidative stress, pH, and temperature), like the bacterial and archaeal hyperthermophiles which contain HARP, and subcellularly in the mitochondria and chloroplasts of eukaryotes.

There are a limited number of functional ribozymes in cells, but of the few that exist all but the ribosome accelerate phosphoryl transfer reactions of various kinds (5), and nearly all are involved in the processing of genetic information. In contrast, cellular metabolic processes are mainly catalyzed by proteins. In the RNA world theory, RNA-based catalysis in early cellular

life predominated, and was eventually replaced by proteins through selection for a variety of properties, including enhanced catalytic efficiency. The retention of an RNA-based RNase P in so many organisms may be due to the similarity in catalytic strategy and efficiency of the RNA-dependent and protein-only forms of RNase P. Consistent with this, the *S. cerevisiae* RNA-dependent RNase P has a value of k_{cat}/K_M that is diffusion-controlled (~10^8 $M^{-1}s^{-1}$) indicating that this is a "perfect" enzyme where the activity is limited by diffusion and evolutionary pressure cannot increase the catalytic efficiency further (6). This suggests that there is little evolutionary pressure to enhance efficiency, thereby leading to cellular retention of these enzymes. In contrast, for other catalytic functions the ribozymes may be less effective leading to more evolutionary pressure to develop better catalysts.

4.4 References

1. Liu, X., Wu, N., Shanmuganathan, A., Klemm, B. P., Howard, M. J., Lim, W. H., Koutmos, M., Fierke, C. A. (2019). Kinetic mechanism of human mitochondrial RNase P. *bioRxiv*, 666792.
2. Reinhard, L., Sridhara, S., & Hällberg, B. M. (2017). The MRPP1/MRPP2 complex is a tRNA-maturation platform in human mitochondria. *Nucleic acids research*, *45*(21), 12469–12480
3. Nickel,A.I., Wäber,N.B., Gößringer,M., Lechner,M., Linne,U., Toth,U., Rossmanith,W. and Hartmann,R.K. (2017) Minimal and RNA-free RNase P in Aquifex aeolicus. *Proc. Natl. Acad. Sci. U. S. A.*, **114**, 11121–11126.
4. Schwarz,T.S., Wäber,N.B., Feyh,R., Weidenbach,K., Schmitz,R.A., Marchfelder,A. and Hartmann,R.K. (2019) Homologs of aquifex aeolicus protein-only RNase P are not the major RNase P activities in the archaea haloferax volcanii and methanosarcina mazei. *IUBMB Life*, **71**, 1109–1116.
5. Lilley,D.M.J. (2011) Mechanisms of RNA catalysis. *Philos. Trans. R. Soc. B Biol. Sci.*, 10.1098/rstb.2011.0132.
6. Hsieh,J., Walker,S.C., Fierke,C.A. and Engelke,D.R. (2009) Pre-tRNA turnover catalyzed by the yeast nuclear RNase P holoenzyme is limited by product release. *RNA*, 10.1261/rna.1309409.